AI Integrated Framework for Intelligent Geospatial Handling and Robust Operation in MultiGIS Applications (AI4MultiGIS)

Topic: Multidimensional Geographic Information Systems (MultiGIS)

AI4MultiGIS

D2.4

Deliverable D.2.4 Pilot Case Study Specification

Work Package	WP2
Delivery Date	M6
Responsible Partner	UVT
Authors	Marc Frincu (UVT)
Contributors	Lucian Parvulescu (UVT) Andrei Ancuta (UVT) Helga Hochbauer (UVT) Maryam Imani (ARU) Maitane Urrutia-Aparicio (UVT) Zhuoqian Wu (ARU)
Distribution	*Project level*
Version	0.20

Deliverable Factsheet

Project Acronym:	AI4MultiGIS
Project Title:	AI integrated framework for intelligent geospatial handling and robust operation in MultiGIS applications
Call:	Chist-Era 2023
Start date:	02/01/2025
Duration:	36 Months

Deliverable Name:	D2.4 Pilot Case Study Specification
Related WP:	WP 2
Due Date:	M6

Editor:	Marc Frincu (UVT)
Contributor(s):	Lucian Parvulescu (UVT) Andrei Ancuta (UVT) Helga Hochbauer (UVT) Maryam Imani (ARU) Maitane Urrutia-Aparicio (UVT) Zhuoqian Wu (ARU)
Reviewer(s):	ARU
Approved by:	All partners

Executive Summary

The AI4MutiGIS project aims to offer an integrated framework for Multidimensional Geographical Information System (MultiGIS) data generation and management, aiming to enhance the overall GIS capabilities for optimised processing chain of MultiGIS applications. This deliverable, D2.4, provides three pilot case studies that will be used to demonstrate the results of the project.

The deliverable details three pilot case studies ranging from sustainable planning and endangered species management to heritage studies:

Pilot 1: *SuDS design and planning through data integration for improved SuDS decision-making in the local region, Essex council*. Implemented by ARU.

Pilot 2: *Detection, tracking, and prediction of the spread of invasive crayfish in Western Romania*. Implemented by UVT.

Pilot 3: *Archaeoastronomical analysis of heritage sites, Al-Hayit and Khaibar, Saudi Arabia*. Implemented UVT.

Each pilot consists of a general description, GAP analysis, ethical requirements, need for MultiGIS, one or more use cases and requirements. These will form the basis for WP5, specifically, Task 5.1 leading to D5.1.

Document History

Version	Date	Author(s)	Comments
0.1	27.01.2025	Marc Frincu (UVT)	Added ToC and the main sections including templates for populating each with information.
0.2	04.02.2025	Marc Frincu (UVT)	Updated ToC based on feedback from the 29.01.2029 project meeting.
0.3	21.02.2025	Helga Hochbauer(UVT)	Added the description (2.2.1), the existing infrastructure (2.2.2) and the engaged users (2.2.4)
0.4	03.03.2025	Helga Hochbauer (UVT)	Uploaded the diagram for 2.2.2 and added the Business needs for 2.2.3
0.5	06.03.2025	Helga Hochbauer (UVT)	Updated the diagram in 2.2.2 and added screenshots, added some bold keywords.
0.6	12.03.2025	Andrei Ancuta (UVT)	Added sections 2.2.5 *Ethical requirements*, 2.2.6 *Need for MultiGIS AI* and 2.2.7 *Use cases*
0.7	13.03.2025	Helga Hochbauer (UVT)	Updated the description (2.2.1) based on the proposal document.
0.8	24.03.2025	Helga Hochbauer (UVT)	Modified the Business needs and added the Future enhancements
0.9	05.04.2025	Maryam Imani (ARU)	Added general information on Pilot Case study 1
0.10	22.04.2025	Helga Hochbauer(UVT)	Data Requirements
0.11	02.06.2025	Andrei Ancuta (UVT)	Functional Process Requirements, Modified 2.2.8 and 2.2.9
0.12	03.06.2025	Maitane Urrutia-Aparicio (UVT)	Added Pilot Case study 3
0.13	03.06.2025	Marc Frincu (UVT)	Reviewed Pilots 2 & 3.
0.14	10.06.2025	Maryam Imani (ARU)	Updated Pilot Case Study 1
0.15	12.06.2025	Andrei Ancuta (UVT)	Modified 2.2.9
0.16	20.06.2025	Zhuoqian Wu (ARU)	Added use case and data requirements on Pilot Case Study 1

Version	Date	Author(s)	Comments
0.17	23/06/2025	Maryam Imani (ARU)	Reviewed & updated Pilot Case Study 1
0.18	29/06/2025	Zhuoqian Wu (ARU)	Modified Pilot Case Study 1
0.19	30/06/2025	Maryam Imani (ARU)	Reviewed & updated Pilot Case Study 1 (following the latest modification)
0.20	01/07/2025	Marc Frincu (UVT)	Added Executive Summary, Table of Figures, List of Acronyms, and References

Table of Contents

List of Figures

List of Tables

Acronyms and Abbreviations

AOI – Area of Interes

AI – Artificial Intelligence

BGS – British Geological Survey

DEM – Digital Elevation Model

GIS – Global Information Systems

MCDM – Multi-criteria Decision-Making

NBS – Nature-Based Solutions

ML – Machine Learning

SCTH – Saudi Commission for Tourism and National Heritage

SDM – Species Distribution Modeling

SuDS – Implementation of Sustainable Drainage Systems

WoC™ – World of Crayfish

1 Introduction

1.1 Purpose and Scope

The aim of this deliverable is to specify the details of three pilot case studies including use case scenarios for data extraction and consolidation from diverse sources and test cases. The requirements will be used for the initiation of the pilot case study operation in WP6. Pilot case study 1 will utilise MultiGIS to enable a more precise SuDS planning and design through enhanced spatial analysis and data integration capabilities. Pilot case study 2 will set up use cases regarding the early detection of invasive crayfish and model its spread by using MultiGIS data and AI. Pilot case study 3 will perform a large scale archaeoastronomical analysis of heritage sites by employing AI on various multispectral satellite images at different resolutions.

1.2 Contribution to other Deliverables

WP5, T5.1, D5.2.

2 Pilot case studies

2.1. Case study 1: SuDS design and planning through data integration for improved SuDS decision-making in the local region, Essex council, UK.

2.1.1 Description

Garden Communities is an emerging urban development initiative in the UK that follows Garden City principles and aims to create sustainable, well-planned developments that provide high-quality housing, green spaces, and modern infrastructure. These communities aim to balance urban growth with environmental considerations, integrating principles of sustainable living, including green infrastructure, biodiversity enhancement, and efficient transport links.

Implementation of Sustainable Drainage Systems (SuDS) and other Nature-Based Solutions (NBS) are at the heart of Garden Communities development in order to reduce flood risk, improve water quality, and enhance biodiversity and create healthier living environments. Traditional drainage systems often struggle to cope with increasing urbanisation and extreme weather events, leading to waterlogging, pollution, and strain on existing infrastructure. SuDS offer a more sustainable approach by mimicking natural hydrological processes, promoting infiltration, storage, and gradual release of water. Their integration into urban planning ensures that cities remain resilient to climate change while meeting the demands of housing market without compromising the environment and providing additional benefits such as green spaces, improved air quality, and enhanced community well-being. Effective SuDS planning is essential for creating liveable, environmentally friendly cities that balance development with ecological preservation.

This case study focuses on part of these Garden Communities in North-East Chelmsford, Essex, UK called Beaulieu and Channels which is a key housing development area in Essex (Figure 2.1.1). Beaulieu will deliver up to 3,600 homes set within 71 hectares of open space including allotments, play facilities, community gardens, sports facilities and reinstated historic estate parkland. Channels is located to the south-west of the Site and comprises up to 750 new homes set within 21 hectares of open space including a Country Park, a series of lakes, skate park, play areas and growing areas.

SuDS have been across Beaulieu and Channels to manage surface runoff. SuDS will be promoted within the Strategy to prevent the wider environment being adversely affected by increased surface water runoff and the increased risk of pollution as a result of the development. This will be achieved through the provision of treatment and storage features including, for example, detention basins, swales, green roofs, permeable paving, etc. to suit the character of the area in which they are positioned.

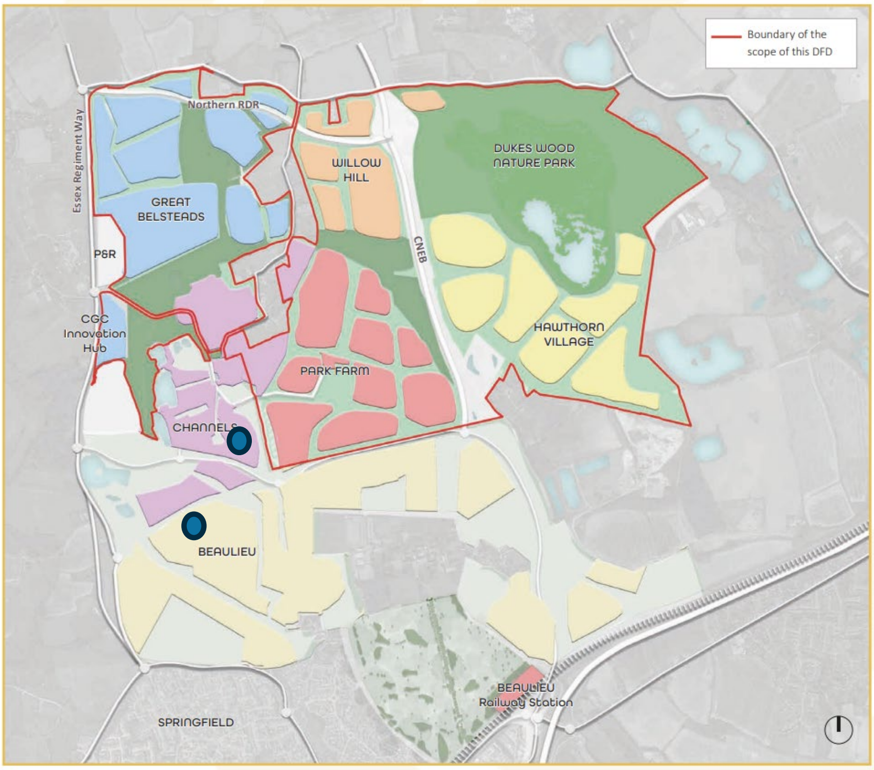


Figure 2.1.1: Chelmsford Garden Community: Beaulieu and Channels (Chelmsford Garden Community Consortium, January 2023)

Generally, designing and planning SuDS using GIS presents several challenges, despite the technology's spatial analysis capabilities. A key issue is the limited integration of dynamic environmental, hydrological, and socio-economic data within GIS platforms, which restricts a holistic understanding of site-specific constraints and opportunities. GIS tools often rely on static datasets and lack real-time or seasonal variability inputs, leading to oversimplified assessments. There is also a gap in interoperability between GIS and hydraulic modelling tools, hindering robust scenario testing and performance prediction.

One of the other key issues in most GIS applications is that they do not adequately incorporate stakeholder priorities, land ownership complexities, or maintenance requirements, which are crucial for SuDS implementation. This is also a significant gap when designing for this case study as per the conversation of the team with local authorities who were responsible for planning permission proposal approval.

As a result, while GIS is useful for initial site screening and visualisation, its current use falls short in supporting the nuanced, multi-criteria decision-making processes needed for effective, context-sensitive SuDS planning.

This project aims to integrate publicly available datasets with local knowledge and stakeholder input to enhance GIS-based planning of SuDS. By embedding multi-criteria decision-making (MCDM) factors at the earliest stages of feasibility assessment, the approach seeks to ensure that SuDS interventions are context-sensitive, inclusive, and aligned with both technical and community priorities. To overcome existing barriers—such as limited data integration, lack of interoperability, and fragmented stakeholder engagement—the project also emphasises improving data accessibility, combining GIS with AI and remote sensing technologies, and developing an effective SuDS GIS Plug-in.

2.1.2 Existing infrastructure and current practice

In SuDS planning and design, publicly available datasets play a crucial role across various stages—from initial site selection to detailed hydraulic modelling. At the *strategic planning stage*, Ordnance Survey historic maps and Google Earth aerial maps are used to better understand the history of the site. Additionally, national datasets such as the Environment Agency's *flood risk maps*, water quality data, and source protection zones are used to understand the extent of site constraints and opportunities. British Geological Survey's *soil and geology data*, and UKCEH *land cover maps* are used within GIS platforms to identify suitable areas for different SuDS components (e.g., infiltration vs. conveyance-based systems). Planners often overlay Ordnance Survey maps, terrain data (e.g., LiDAR), and flood extents to assess surface water flow paths, potential flood hotspots, and topographical constraints. These layers help screen and prioritise intervention zones, especially in urban retrofit scenarios where spatial constraints are significant.

During the *design and feasibility stage*, datasets become more site-specific and technical. UK SuDS Tools such as HR Wallingford web-based tools, assist in the design and evaluation of SuDS schemes. They include tools for estimating greenfield runoff rates, surface water storage volumes, and infiltration volumes. Local authority *drainage asset maps*, *groundwater* vulnerability data, and *rainfall and climate data* intensity records from the Met Office are used to inform calculations of runoff volume, infiltration potential, and pollution risk. Tools like SCIMAP (sediment sized mapping) and the SuDS Tool utilise such inputs to guide decision-making on the type and sizing of SuDS features. Additionally, socio-environmental datasets from sources like DEFRA's MAGIC Map or local planning constraints are integrated to ensure designs comply with environmental regulations (e.g., protected habitats) and deliver wider co-benefits.

Drawing on this, the following sources or infrastructure/platforms are used:

1. **UK SuDS Tools such as HR Wallingford for estimating greenfield runoff rates, surface water storage** volumes, and infiltration volumes.
2. **British Geological Survey (BGS) Infiltration SuDS Map** for data on subsurface permeability, depth to groundwater, and geological constraints.
3. **Environment Agency Data** for flood risk maps, water quality data, and source protection zones.
4. **Local Authority Data/Information** for land use data, planning constraints, and historical flooding events.
5. **Ordnance Survey Data** for topographic maps, digital terrain models, and boundary data.
6. **Met Office Data** for rainfall data, climate projections, and extreme weather event records

Despite the wide availability of data, integration across datasets and among stakeholders remains fragmented, highlighting the need for more seamless and interoperable GIS-based workflows to fully realize the multifunctional potential of SuDS. Currently, some Local Authorities in the UK (e.g., Greater Manchester, London) have developed 'SuDS opportunity maps' along with a password-protected web tool that combine multiple data layers. London SuDS Opportunity Mapping Tool[1] has been also developed to aid Local

---

[1] https://www.arcgis.com/apps/webappviewer/index.html?id=a386e0d5307e4db1a30e95cd39399554#:~:text=When%20using%20SuDS%20Opportunity%20Map,used%20as%20a%20decision%20tool (accessed July 1, 2025)

Authorities and other Risk Management Authorities in understanding the potential for SuDS within any given area, providing approximate volumes of surface water that require managing and assigning approximate costs to the SuDS features. The challenges can be categorised in four main groups:

1. Data-Related Challenges

- **Data Availability and Quality:** High-resolution spatial data (e.g., land use, soil permeability, drainage networks) may be incomplete, outdated, or inconsistent across different sources.
- **Data Standardisation and Compatibility:** Different agencies and stakeholders may use varying formats and classification systems, making data integration difficult.
- **Access to Proprietary Data:** Some key datasets, such as detailed hydrological models or urban drainage networks, may be unavailable, restricted or costly to obtain.
- **Lack of Real-Time Data:** SuDS performance and urban water dynamics require real-time or high-frequency monitoring data, which is often unavailable.

2. Technical Challenges

- **Modelling Complexity:** Accurately simulating SuDS performance requires integrating GIS with hydrological and hydraulic models, which can be complex, sometimes costly and computationally intensive.
- **Scale and Resolution Issues:** SuDS planning requires both large-scale regional analysis and detailed site-specific assessments, which can be challenging to balance in GIS-based workflows.

3. Methodological Challenges

- **Multi-Criteria Decision Making (MCDM):** SuDS planning involves multiple factors (e.g., flood risk, pollution control, biodiversity, cost), and GIS-based MCDM approaches may require complex weighting and stakeholder input, for example, trade-offs between water quantity vs. land use efficiency, biodiversity and amenity vs. maintenance burden, retrofit feasibility vs. performance etc.
- **Uncertainty and Sensitivity Analysis:** GIS models rely on assumptions and input parameters that introduce uncertainties, which must be accounted for in decision-making.
- **Climate Adaptation Considerations:** GIS tools must incorporate future climate scenarios (e.g., rainfall intensity, temperature changes), but reliable projections at the local scale can be difficult to obtain.

4. Practical and Institutional Challenges

- **Stakeholder Collaboration and GIS Expertise:** Effective use of GIS for SuDS requires skilled professionals and collaboration between planners, engineers, and environmental scientists, which may not always be available.
- **Decision-Making Complexity:** GIS outputs must be interpretable for non-technical stakeholders, requiring user-friendly visualisation and communication strategies.

- **Policy and Regulatory Alignment:** GIS tools should align with planning regulations and sustainability policies, but a lack of standardized methodologies can lead to inconsistent applications

2.1.3 Future enhancements

While addressing all the aforementioned challenges is beyond the scope and feasibility of this project, it is important to maintain a vision for potential future engagements where needed. This study primarily leverages publicly available maps, datasets, local knowledge, and QGIS to address some of the identified limitations, aiming to enhance SuDS planning and decision-making while managing some key uncertainties associated with SuDS implementation.

The following improvements are expected:

FE1. Integration and Standardisation of Data - Integrate geospatial layers describing different data needed as well as local knowledge for SuDS planning and design.

FE2. Rapid Performance Estimation at SuDS Feasibility Study stage – Development of the site Surrogate Model enables estimation of key performance indicators informing decision makers at feasibility, and conceptualisation stage.

FE3. Improved SuDS Muli-Criteria Decision Making – the above two improvements will lead to improved decision making at planning stage for SuDS.

FE4 – Improved accessibility - A publicly available Plug-and-Play for SuDS feasibility study.

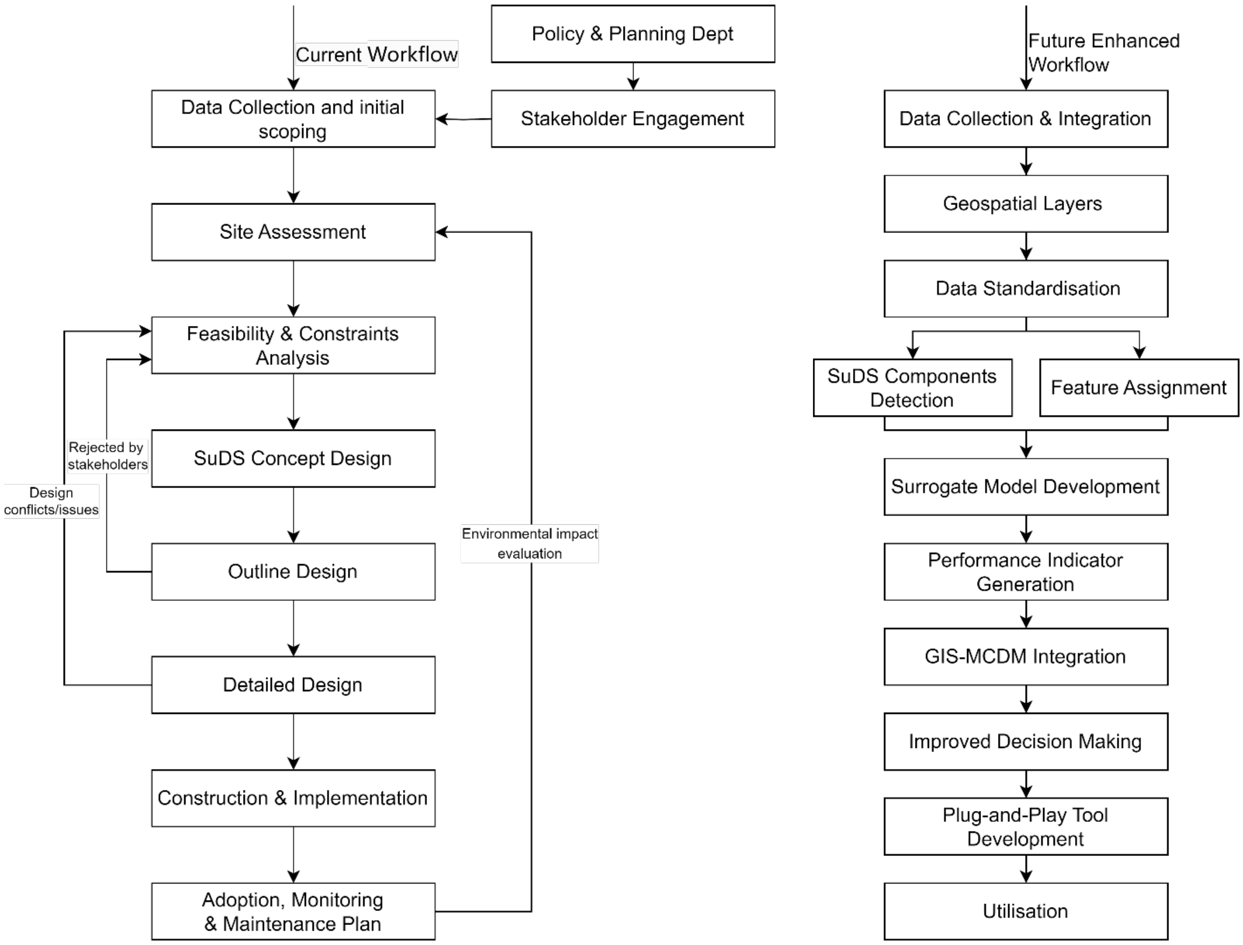


Figure 2.1.2. Current SuDS planning and design workflow and the proposed future enhancements

2.1.4 Business needs

Case Study 1 has already been constructed and is currently occupied by residents. It was previously nominated for the UK's Best SuDS Design Award, underscoring its recognition as a benchmark for best practice in sustainable drainage. Owing to its comprehensive design and demonstrated success in meeting key SuDS performance criteria, the site serves as a valuable reference for future SuDS planning. By analysing and replicating its design rationale—supported by local knowledge and contextual factors—we aim to develop a decision-support tool that is both adaptable and replicable across diverse locations. This tool will guide future SuDS implementation through a structured MCDM approach. The primary objectives are: To identify appropriate SuDS options (i.e., the SuDS management train); To determine optimal locations for their implementation.

BN11: Comprehensive Characterisation, Standardisation and Integration of the Site Data

To enable informed and efficient SuDS planning, the system must support the comprehensive characterisation of SuDS sites such as Beaulieu and Channels. This includes ingesting diverse spatial data layers (aerial images, DEMs, land cover maps), detailed site design documents, performance metrics, and environmental metadata. The system should digitise, georeference, and integrate these heterogeneous datasets to generate detailed spatial representations and performance analyses.

Input Data

- Spatial layers: aerial images, Digital Elevation Models (DEM), land cover maps
- Site design documents detailing SuDS components (e.g., swales, bioretention areas)
- Performance data: runoff reduction, water quality improvements, infiltration rates
- Metadata: soil characteristics, hydrology, local climate conditions

Processing & Integration

- Digitise and georeference design elements within GIS
- Integrate performance data with spatial layers using attribute joins or spatial overlays
- Analyse spatial patterns to identify key factors contributing to site success (e.g., slope, proximity to water)

Outputs & User Interaction

- Detailed spatial "signature" of the SuDS site as layered GIS projects
- Interactive maps allowing engineers and planners to query performance metrics and design attributes
- Summary reports highlighting key success factors for communication and planning

Technical Requirements

- GIS software (e.g., QGIS)

- Remote sensing tools (e.g., Google Earth Engine)
- Data processing and spatial analysis capabilities

BN12: SuDS-AI MetaPlanner – Interactive maps that allow querying individual SuDS features Surrogate Model for SuDS Feasibility Analysis and Decision Making

To enhance the scalability and objectivity of SuDS planning, the system must incorporate AI/ML models trained on spatially labelled data to predict SuDS type and location suitability. This includes preprocessing data layers, training classification models (e.g., Random Forest), and integrating suitability predictions within GIS for visualization and interactive threshold adjustments. The SuDS-AI MetaPlanner outputs should feed into a MCDM framework (such as AHP) and be validated against real-world data to ensure accuracy.

Input Data

- Labelled datasets distinguishing known SuDS sites as “suitable” and others as “not suitable”
- Extracted values from GIS layers representing criteria for suitability (e.g., slope, soil type)

Processing & Integration

- Preprocess data (normalization, missing value handling, feature engineering)
- Split data into training and testing datasets
- Train machine learning models (e.g., Random Forest) to predict suitability
- Evaluate model accuracy (confusion matrix, ROC curves)
- Integrate AI predictions with GIS data

Outputs & User Interaction

- Trained AI model that produces suitability scores or probability maps
- Visual suitability maps in GIS for easy interpretation
- User options to adjust model thresholds or retrain as needed
- Support for MCDM to incorporate expert inputs and stakeholder priorities

Technical Requirements

- Machine learning libraries (e.g., scikit-learn, TensorFlow, Keras)
- Computing environment for model training (local or cloud)
- Scripting proficiency (e.g., Python, R)
- QGIS for displaying AI outputs (e.g., Python GIS libraries…)

2.1.5 Engaged users

The primary users of an AI4MultigIS application for SuDS planning are urban planners, water resource engineers, environmental consultants, and local authority decision-makers. These professionals are directly involved in the site selection, design, and evaluation of sustainable drainage systems and require evidence-based, spatially informed tools to support integrated water management. By engaging these users early in the development process—through interviews, surveys, and participatory workshops—the system can be aligned with practical needs, local constraints, and regulatory frameworks, ensuring that it supports real-world decision-making.

In addition to technical professionals, secondary users may include stakeholders from environmental NGOs, developers, infrastructure agencies, and community groups interested in urban resilience and green infrastructure. These users benefit from accessible interfaces, scenario modelling, and visual outputs that communicate the benefits and trade-offs of SuDS strategies. Engaging these broader stakeholders helps build trust, supports transparent planning, and increases the likelihood of successful implementation and long-term maintenance of SuDS features in the urban landscape.

The integration of AI, GIS, and MCDM in SuDS planning offers researchers a powerful, interdisciplinary platform for advancing urban sustainability, spatial analysis, and decision science. By using real-world data and a validated SuDS site, academics can test novel models, evaluate spatial criteria, and publish applied case studies with real impact. An open-source implementation significantly enhances this value—enabling reproducibility, collaborative development, and adaptability across diverse urban contexts, while lowering barriers for researchers in resource-constrained settings.

2.1.6 Ethical requirements

- **Data Privacy and Consent**

  Use of spatial data may include sensitive infrastructure, land ownership, or demographic information. Hence, all personal or sensitive data are anonymised and/or appropriate permissions for any site-specific or user-contributed data (e.g., from local councils or residents) will be obtained (if needed).

- **Algorithmic Transparency and Bias**

  Machine learning models may carry biases based on training data, leading to skewed results or unsuitable recommendations. Model transparency and explainability (e.g., via SHAP values or feature importance) will be maintained. Data sources and model decisions are documented. Peer review or open-source publication of model code will be carried out.

- **Stakeholder Engagement and Informed Consent**

  AI-GIS systems may marginalise human judgment or exclude local knowledge if not designed inclusively. Key stakeholders are involved throughout development. Outputs will be made interpretable and provide channels for feedback or dissent.

- **Open Access and Reproducibility**

  Where possible, we use open-source tools and publish our methodology/code to promote reproducibility and collaborative research.

2.1.7 Need for MultiGIS AI

The integration of MultiGIS AI into SuDS planning addresses technical and planning complexities, significant skill demands, and the need for transdisciplinary coordination from concept to implementation. Below, we outline the key reasons why incorporating AI and its capabilities can significantly enhance decision-making in SuDS. As SuDS increasingly become a mainstream approach for nature-informed planning, AI offers powerful tools to address environmental, ecosystem, and social challenges more effectively.

1. To Handle Complex, Multi-Criteria Decision-Making

SuDS planning involves balancing hydrological, environmental, social, and spatial factors. Traditional GIS-MCDM tools require manual weighting and analysis, which can be subjective and time-consuming. AI enhances this process by learning optimal patterns from past data (e.g., your reference SuDS site), reducing bias and improving decision quality.

2. To Leverage Existing Best-Practice SuDS Sites

By training AI models on successful SuDS implementations, such as the best practice case study, the underlying spatial and functional logic can be extracted and replicated. This allows scaling best practices across new sites without reinventing the analysis for each project.

3. To Improve Spatial Suitability Predictions

GIS-based overlays are static and depend heavily on expert-defined weights. AI models (e.g., Random Forests or XGBoost) can dynamically learn from multiple spatial layers—such as slope, soil, land use, and drainage—to predict where SuDS components are most likely to perform well.

4. To Enhance Transparency and Reproducibility

Using open-source AI tools within a GIS framework encourages transparent, data-driven decision-making. Unlike black-box decisions or consultant reports, AI workflows can be documented, shared, tested, and improved by others, strengthening the credibility and scientific rigor of SuDS planning.

5. To Support Scenario Testing and Adaptability

AI models can quickly re-run predictions under different climate, land-use, or stakeholder-driven scenarios. This flexibility allows planners to test multiple SuDS configurations, optimise placement, and future-proof designs against environmental uncertainty.

6. To Democratize SuDS Planning Through Smart Tools

With AI-integrated MultiGIS platforms, even non-experts can explore SuDS suitability and planning options through intuitive, map-based interfaces. This makes decision-support tools more accessible to Local Authorities, different stakeholders, not just technical consultants.

2.1.8 Use cases

ID:	**SuDS_UC_01**
Title:	Characterise and Integrate SuDS Site Data for Spatial Analysis

Description:	This use case involves applying AI techniques to **detect, classify and characterise existing SuDS components within the site drainage network (e.g., in Beaulieu and surrounding channels).** Using design documents and GIS data, the system identifies the type and location of SuDS elements and integrates their associated performance results with spatial layers via attribute joins and overlays. The resulting multi-layered interactive map presents the spatial "signature" of these SuDS components, allowing users to query design attributes and performance metrics. This spatial reference model can be reused to support SuDS recognition and planning in similar urban areas.
Primary Actor:	Developers, urban planners, highway and drainage engineers, environmental consultants, ecologists, public health experts
Preconditions:	• Geospatial data layers should include topography, Digital Elevation Models (DEMs), land cover and land use classifications, hydrological networks (e.g., rivers, drainage paths), geological formations, soil types, flood risk zones, climate data (e.g., rainfall intensity and frequency), and other locally relevant environmental and infrastructural datasets. • **Publicly available site design documents**, such as SuDS design manuals, council guidelines, or planning documents. • **Project-specific technical reports (optional)**, including detailed engineering designs or as-built documentation, which may offer more precise information on the location and type of SuDS features but may be confidential or restricted in access. • AI model capable of performing semantic segmentation of SuDS types (or SuDS train) from satellite imagery. • GIS environment prepared for spatial data integration, analysis, and interactive map generation (e.g., QGIS with required plugins).
Postconditions:	• A semantically segmented SuDS land cover dataset for SuDS sites such as Beaulieu and Channels. • Visualization of the spatial "signature" of the SuDS site. • Interactive maps that allow querying individual SuDS features. • Exportable summary reports that highlight spatial patterns and key success factors.
Business needs:	• Ingests and harmonises diverse spatial datasets, including aerial imagery, digital elevation models (DEMs), and land cover maps, to provide a unified geospatial foundation for site understanding. AI-driven semantic segmentation tailored to SuDS enables design-specific classification for planning and analysis. • Links performance metrics to SuDS features via attribute joins or spatial overlays, enabling performance-informed spatial reasoning.

	- Strengthen integration between GIS platforms and external hydraulic and hydrological modelling tools, enabling efficient data exchange. - Develop a multi-layered, interoperable spatial platform supporting collaborative, informed, and context-sensitive SuDS planning.
Main Success Scenario:	1. User uploads and imports required spatial datasets, including DEM, satellite/ aerial imagery, land cover maps, soil and hydrology layers and original SuDS design documents. 2. System automatically detects, segments, and georeferences the SuDS design elements and overlays them onto base maps. 3. SuDS design elements are digitised as vector layers with unique identifiers. 4. The system integrates built-in performance datasets with spatial features using attribute joins or spatial overlays. 5. Platform conducts spatial pattern analysis and generates correlation layers between performance outcomes and spatial variables such as slope, land use, and soil type. 6. Users query and review the performance metrics and design attributes and label spatial units as “suitable” or “not suitable” for different SuDS types. 7. The platform filters the relevant zones and features and outputs a packaged QGIS project, labelled training dataset, and a summary report of spatial success factors.
Extensions:	- **Low-Resolution Imagery Detected:** If imported aerial or satellite images are below the required resolution threshold, the system flags the issue and informs the user that segmentation accuracy may be influenced and recommends uploading higher-resolution imagery. - **Incomplete Or Outdated Spatial Layers:** If the essential GIS layers (such as DEM, soil, and land cover) uploaded by the user do not cover the target area or are beyond the valid time limit, the user will be alerted to replace the data or use interpolated data - **Label Imbalance in Training Dataset:** If the majority of the spatial units is labelled as " not suitable", the system will alert the user that the training categories are imbalanced, and the data requires evaluation. Positive samples will be incorporated if required to ensure that the training data is balanced.
Frequency of Use:	This tool has the potential to be adopted by local authorities, design engineers, and developers during the SuDS feasibility study stage. Its frequency of use will depend on the scale and number of projects that choose to implement it. However, the primary goal is to apply and refine the tool in collaboration with Chelmsford City Council—where the case study is located and who are a key partner in the project’s development and validation as well as deployment eventually.

Status:	In development
Owner:	ARU
Priority:	HIGH

ID:	**SuDS_UC_02**
Title:	Predict and Visualise SuDS Suitability Using AI-Driven Modelling
Description:	This use case describes the process of training an AI model (e.g., random forest) to predict the suitability of SuDS sites using spatially labelled data. The AI model outputs probabilities or suitability scores that can be visualized in GIS. The model can be fine-tuned to customize the output through user interaction with the model. The results are also integrated into a multidimensional decision model (MCDM) framework (e.g., analytic hierarchy process (AHP)) to gather expert opinions to achieve informed decisions across the gap between theory and practical application. The model supports scalable, objective, and data-driven SuDS planning.
Primary Actor:	Developers, urban planners, hydrologists, environmental engineers, and decision-makers involved in SuDS planning and site selection.
Preconditions:	• Access to pre-labelled spatial training data indicating suitable/unsuitable SuDS spatial units. • Extracted values from GIS layers representing key suitability criteria (e.g., slope, soil type, land use, flood risk) are available and linked to spatial units. • Functional computing environment with installed ML libraries (e.g., scikit-learn) and GIS tools (e.g., QGIS). • Access to ground-truth data for validation (e.g., Channels)
Postconditions:	• A trained AI model capable of predicting suitability across new regions. • A suitability map embedded in GIS showing predicted SuDS planning. • The trained AI model outputs evaluation metrics (e.g., accuracy, F1-score)

	• An interactive interface where users can adjust thresholds and retrain the model as needed. • A multi-criteria decision-making framework incorporating expert knowledge into the final suitability analysis.
Business needs:	• AI-driven analysis delivers large-scale regional analysis and detailed site-specific assessments in minutes, removing the traditional “scale vs resolution” trade-off. • Seamless linkage of AI outputs to AHP and other MCDM engines unifies flood risk, pollution control, biodiversity, and life-cycle cost into one consistent score. • Interactive controls for weights and thresholds let stakeholders observe ranking shifts, resulting in decisions that are both transparent and customised. • Built-in diagnostics such as confusion matrices, ROC curves, and Monte Carlo sensitivity quantify uncertainty, improve model robustness, and make AI models interpretable before plans are finalized • A visual, guided interface lowers the GIS/ML barrier so planners, engineers, and ecologists can contribute without coding expertise. • Case-validated models and weight presets are portable to new communities, creating auditable standards while cutting future study time and cost.
Main Success Scenario:	1. Users select a labelled dataset, containing spatial units classified as "suitable" or "not suitable" for SuDS, along with associated environmental variables (e.g., slope, land use, soil type). 2. System preprocesses the input data, including normalization, missing value handling, and feature engineering to prepare the dataset for training. 3. User initiates model training, selecting an AI algorithm such as Random Forest. 4. Model conducts the training and reports the evaluation metrics (e.g., confusion matrix, F1-score, ROC curve). 5. Users review the results to assess model reliability. 6. System generates a suitability prediction map, visualizing suitability scores or probability maps spatially. 7. Users view the prediction outcomes and choose to adjust the thresholds to fine-tune the model output or retrain the model. 8. Users launch the MCDM module, identify the criteria of interest, and allocate weights to reflect their relative importance. 9. System performs weighted raster overlay, generating an expert-informed suitability map that reflects both model output and stakeholder priorities. 10. Users export final predicted maps and evaluation reports, summarizing spatial layers, score distributions, and MCDM rankings.

Extensions:	- **Model Performance Below Threshold:** If the results of the evaluation metrics are outside the acceptable range, such as the accuracy is too low, the system will mark the problem and prompt training parameter adjustment or feature revision. - **Processing Timeouts or Hardware Limits:** If model training exceeds the computational limit, the task will be interrupted, and the user will be notified.
Frequency of Use:	Depends on the research or the initial plan design
Status:	In development
Owner:	ARU
Priority:	MEDIUM

2.1.9 Functional requirements

Data requirements

AI-driven SuDS suitability modelling and planning system collects and integrates six categories of data.

Entity	**Definition**
Spatial Layer	A collection of spatial data representing a particular class or type of real-world entities in a particular area.
SuDS Component	A suite of components working in different ways to manage flows, volumes, water quality and providing amenity and biodiversity benefits.
Performance Metric	Indicators to quantify SuDS performance include including runoff reduction, water quality improvements, infiltration rates.
Environmental Context	Metadata about soil, hydrology, and climate relevant to each site.
Cost Profile	Initial and maintenance costs of different SuDS components.
Planning Constraint	Regulatory, zoning, or flood risk boundaries affecting site suitability.

Attribute Definitions

Spatial Layer

- Layer ID: Unique identifier for each GIS layer.
- Type: satellite/ aerial images, DEM, land cover maps.
- Source: Copernicus, OS MasterMap, Sentinel-2, Landsat-8.
- Resolution: Spatial resolution in meters.

SuDS Component

- Component ID: Unique identifier for each SuDS feature.
- Type: the SuDS types outlined in UK SuDS Manual[2].
- Reference Document: Link to the design file (PDF/CAD).

Performance Metric

- Metric ID: Unique identifier for each metric.
- Component ID: Associated SuDS feature.
- Type: water quantity management (e.g., runoff volume reduction, storage capacity maximisation, etc), water quality improvement (e.g., % of vegetation cover, ratio of SuDS treatment area to contributing impervious area, soil infiltration rates etc), biodiversity and ecology (e.g., area, number and variety of plants etc), amenity and social value (e.g., availability of SuDS features).
- Source: Monitoring report, SuDS manual and UK SuDS Tools.

Environmental Context

- Context ID: Unique identifier.
- Type: Soil Type, Hydrologic Group, Climate Zone, Rainfall Event.
- Source: British Geological Survey (BGS), Met Office, Environment Agency.

Cost Profile

- Cost ID: Identifier for the SuDS element cost record.
- Component Type: Associated SuDS type.
- TotalCost: Capital and maintenance cost (e.g., £/m^2)
- Cost Source: Environment Agency.

Planning Constraint

- Constraint ID: Unique identifier.
- Type: Flood zone, protected habitat, source protection zone, planning boundary.
- Source: Environment Agency, local authority.

---

[2] The SuDS Manual (C753) - https://www.ciria.org/CIRIA/CIRIA/Item_Detail.aspx?iProductCode=C753

Data Requirements Summary

- **DR1:** Collect and digitise basic spatial data to complete spatiotemporal registration and geographic coordinate unification, which will serve as the base map for SuDS modelling.
- **DR2:** Digitise SuDS design features and integrate them with geospatial layers to achieve spatial positioning and attribute attachment
- **DR3:** Prepare SuDS performance metrics and cost profiles to embed into the system and link to related SuDS components

  DR4: Collate and integrate environmental data and planning constraints to provide a basis for subsequent suitability analysis and decision support.

Functional process requirements

AI integration with MultiGIS for better SuDS planning is a multi-stage process. The key workflow is as follows:

FR1. SuDS Site Characterisation

- Import and manage spatial datasets (DEM, aerial imagery, land cover, etc.)
- Digitise and georeference SuDS design elements
- Integrate performance data with spatial features
- Analyse spatial patterns to identify success factors

FR2. Training Data Preparation

- Compile and standardise environmental and urban GIS layers
- Mask irrelevant zones and extract relevant features
- Label spatial units as “suitable” or “not suitable”
- Create an interactive, queryable GIS database

FR3. AI/ML-Based Suitability Modelling

- Preprocess input data (e.g., normalization, missing values)
- Train classification/regression models (e.g., Random Forest)
- Generate suitability prediction maps
- Provide tools to retrain or adjust thresholds interactively

FR4. Multi-Criteria Decision Making (AHP)

- Input expert judgments via pairwise comparisons
- Calculate criterion weights and consistency ratios
- Perform weighted raster overlays in GIS
- Output expert-informed suitability maps

FR5. Model Validation

- Compare predictions to actual validation sites using metrics (e.g., confusion matrix, F1-score)
- Perform sensitivity analysis with alternative scenarios

FR6. Decision Support & Planning

- Integrate final suitability maps with planning constraints and stakeholder input
- Develop and compare planning scenarios
- Visualise and interact with spatial outputs in QGIS
- Export data and reports for use in policy and design

2. 2. Case study 2: Detection, tracking, and prediction of the spread of invasive crayfish in Western Romania.

2.2.1 Description

This case study focuses on protecting biodiversity and managing **ecological impacts** by leveraging the connection between technology and biological facts. More exactly it consists of gathering the necessary data about the species of invasive crayfish in Western Romania and its usage and analysis that can lead to more precise future predictions.

Invasive species carry diseases that kill native species and impact local economies such small rural enterprises. The impact can expand to the entire supply chain, with restaurants and supermarkets lacking in turn raw materials. The spread of these species driven by anthropic, and climate affects the natural balance of local ecosystems. Migration scenarios require the integration of multidimensional and multitemporal data.

An optimal approach to address the problem is based on an interdisciplinary expertise, where multiple fields of science work together to create innovative solutions for ecological challenges. This integration enables efficient data collection, analysis, and decision-making, ensuring that conservation efforts are both scientifically informed and technologically assisted.

2.2.2 Existing infrastructure and current practice

At this moment, the main available resources regarding the detection, tracking, and prediction of the spread of invasive crayfish in Romania is the portal **World of Crayfish™ (WoC™)**[3] (Ion et al. 2024), which attempts to generate crucial information for people who want to delve deeper into this subject. **WoC™** offers an environment enriched with experts and their research data, designed to be utilized following the highest standards and advanced through state-of-the-art technologies.

Positioned at the intersection of Digital, Green, and Nature domains, **WoC™** aspires to redefine the concept and utility of biodiversity databases. Existing biodiversity databases, often driven by volume and heavily reliant on citizen science, provide large datasets valuable for early alerts on invasive species. However, this focus introduces **challenges**: (1) an imbalance of data skewed toward charismatic species or specific funding priorities and (2) potential inaccuracies due to non-expert contributions. **WoC™** leverages interdisciplinarity to tackle these challenges, integrating biology, GIS, statistics, and IT into a seamless platform. Combining expert-curated data with innovative technologies, it ensures high-standard usability and advanced data processing for research and conservation.

Currently, the platform offers **limited functionalities**. It presents distribution data on a global map and operates using **QGIS Online** (Figure 2.2.1). Users can register, and the administrator reviews and either approves or rejects their request. Once approved, a user gains access to the **exact geographic location of each record** (Figure 2.2.2).

---

[3] https://www.world.crayfish.ro (accessed July 1, 2025)

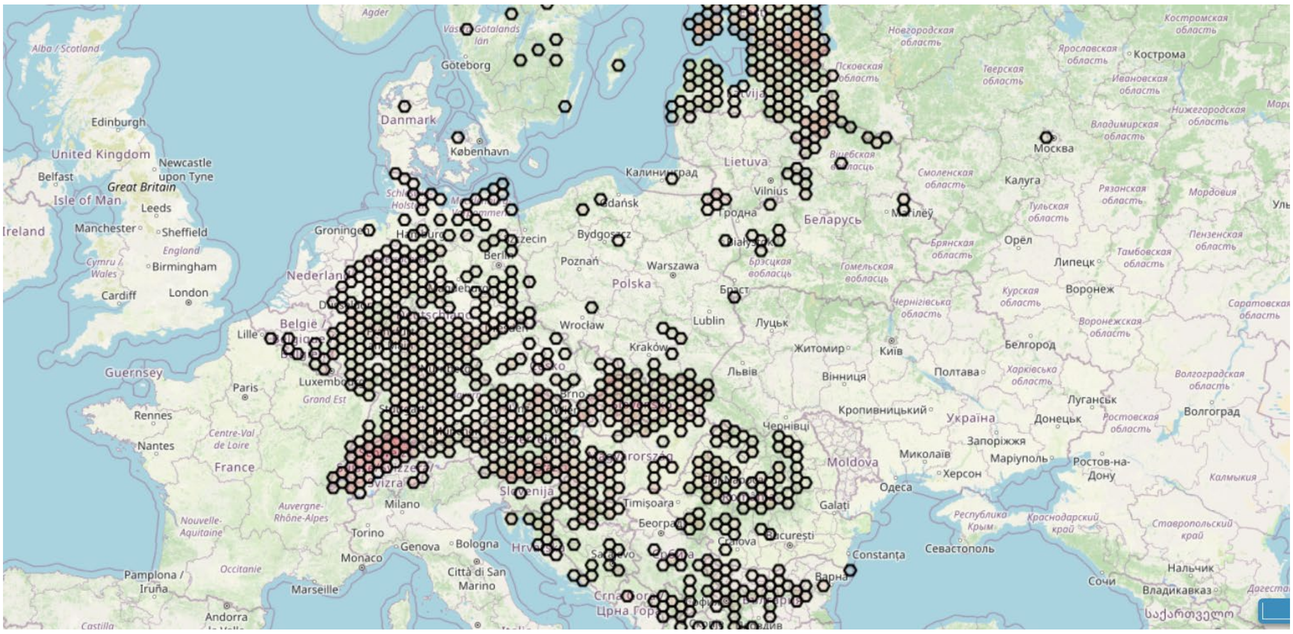
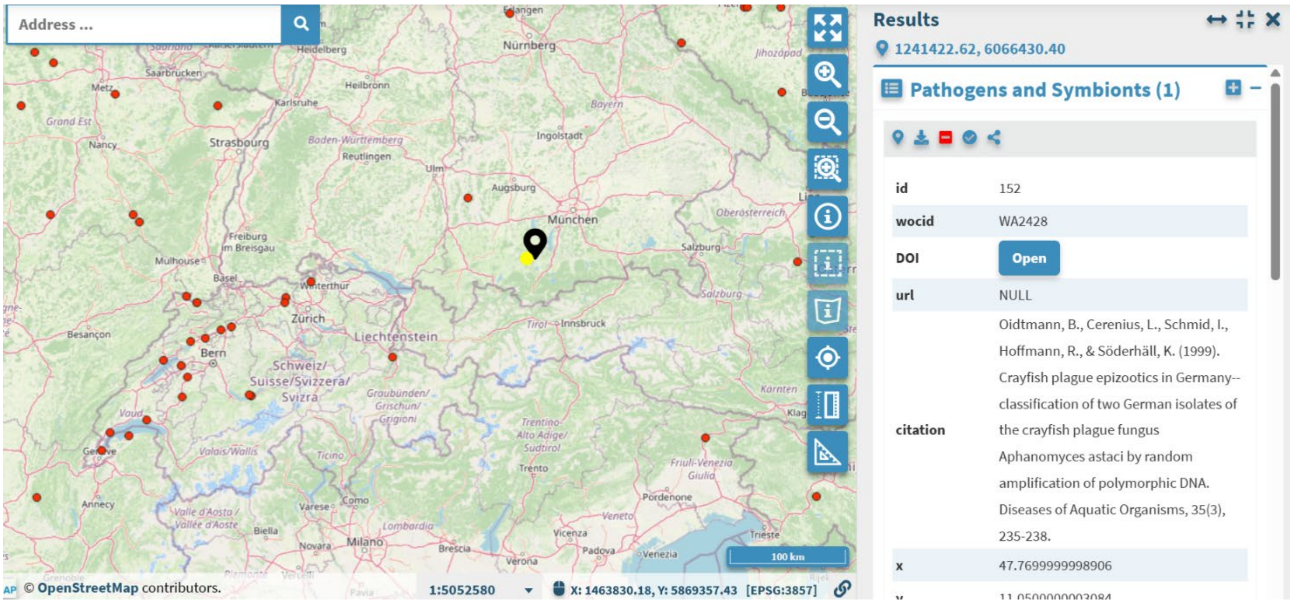


Figure 2.2.1. Overview of the WoC™ map.

Unregistered users can view the distribution through hexagonal grids (Figure 2.2.1) that obscure the precise locations. This feature is intended to protect data related to crayfish species that are vulnerable to exploitation.



X: 1463830.18, Y: 5869357.43 [EPSG:3857]

Results

1241422.62, 6066430.40

Pathogens and Symbionts (1)

id	152
wocid	WA2428
DOI	Open
url	NULL
citation	Oidtmann, B., Cerenius, L., Schmid, I., Hoffmann, R., & Söderhäll, K. (1999). Crayfish plague epizootics in Germany--classification of two German isolates of the crayfish plague fungus Aphanomyces astaci by random amplification of polymorphic DNA. Diseases of Aquatic Organisms, 35(3), 235-238.
x	47.7699999998906
y	11.0500000003084

Figure 2.2.2. Specific details on an existing entry in WoC™.

WoC™ already hosts over 106,000+ records representing more than 400 crayfish species worldwide. Each record has a unique ID, making **WoC™** a potential open data repository with accession codes.

2.2.3 Future enhancements

The necessary future additions to the platform include layers of geospatial data that describe aquatic habitats, an improved user interface, and analysis tools based on existing data (Figure 2.2.3)

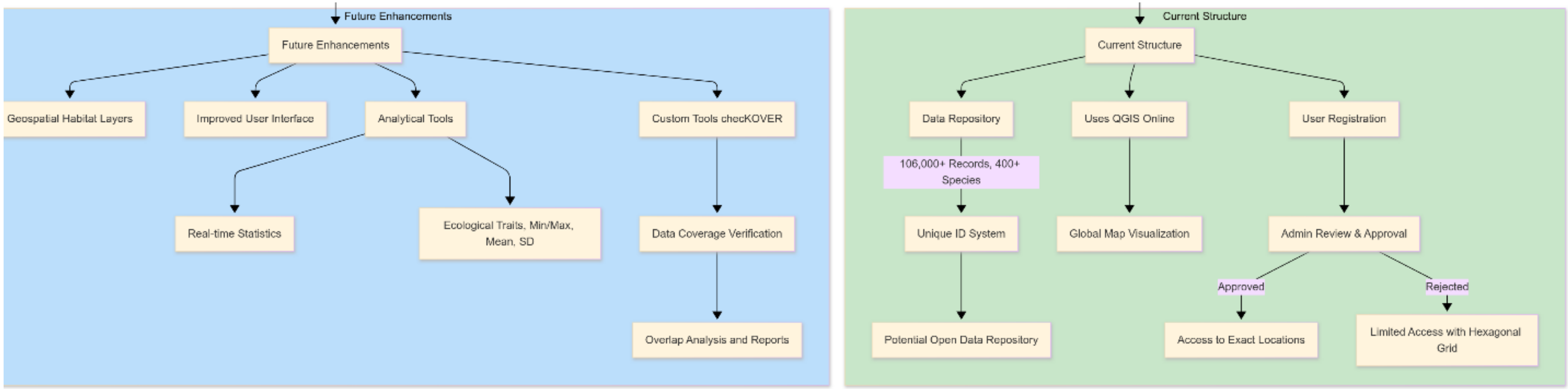


Figure 2.2.3: WoC™ current structure and future enhancements.

FE1. Integration of Habitat Data. Integrate geospatial layers describing aquatic habitats to provide a more comprehensive view of crayfish ecosystems.

FE2. Offline Functionality. Guarantee accessibility and facilitate data contribution during fieldwork, even in the absence of an internet connection.

FE3. Real-Time Data and Reporting. Access to real-time data and automated reports with statistics on species distribution, health, and ecological trends.

FE4. Data Protection and Security. Protect sensitive species data from misuse or exploitation while still providing valuable insights for researchers and conservationists.

FE5. Improved User interface. Building a more accessible user interface that includes the new custom tools.

2.2.4 Business needs

BN21. Predicting Invasive Crayfish Species Using WoC™

WoC™ can be a valuable tool for predicting biological invasions by integrating species occurrence data with geospatial river network variables. The required workflow involves:

- **Input Data:**
    - Existing crayfish records in WoC™.
    - Hydrological and environmental variables from structured geospatial datasets (Domisch et al. 2015), available at a 1x1 km resolution. These datasets include climatic, soil, land use, and hydrological parameters.
- **Processing & Integration:**
    - Linking crayfish records to the corresponding river network data to generate a structured training dataset for **Species Distribution Modeling (SDM)** without real absence data.
    - SDM will identify environmental conditions associated with known occurrences and extrapolate them across selected hydrological networks (country, continent, or custom region).
- **Output & User Interaction:**
    - Users will be able to generate predictive invasion risk maps online or offline.
    - The platform will visualize areas with high invasion risk, aiding in prevention and management strategies.
- **Technical Requirements:**
    - Integration of river network geospatial data into WoC™.
    - Implementation of user-friendly modeling tools accessible via the WoC™ interface.

BN22. CheckOVER – A Tool for Overlap Analysis Between Species

A user-friendly tool designed to assess spatial overlap between two or more species, facilitating insights into competition, co-occurrence, or potential invasive interactions.

- **Input Data:**
    - WoC™'s existing occurrence datasets in both **hexagonal grid format (16 km edge)** and **point data**.

- **Processing & Integration:**
    - Implementation of an interactive tool to compute and visualize species distribution overlaps.
    - Comparison metrics for both point data and hexagonal grids.
- **Output & User Interaction:**
    - Users can select species and receive spatial overlap reports.
    - Visualization of co-occurrence zones through interactive maps.
- **Technical Requirements:**
    - Development of an integrated tool within WoC™ to handle and compute spatial overlap statistics.

Both business needs significantly expand WoC™'s functionality, making it a powerful resource for invasion biology, conservation planning, and ecosystem management.

2.2.5 Engaged users

The key beneficiaries of this case study are **scientists** specializing in biodiversity, ecology, and GIS. In addition, **political decision-makers**, and **governmental authorities** alongside **environmental conservation NGOs** can access and use the scientific data provided by this study in the development of environmental strategies and the monitoring of invasive species.

Decision-makers can leverage the developed analytical tools to assess the spread of invasive species, identify conservation priorities, and develop policies grounded in real-time ecological data. NGOs can utilize our tools for various reasons such as tracking species population trends to inform sustainable environmental strategies or supporting habitat protection initiatives.

In **academia**, the tools and the scientific information available in this case study may serve as an educational resource for students and professors, offering access to a comprehensive dataset and analytical tools that enhance ecological research and **GIS-based studies**.

2.2.6 Ethical requirements

This pilot case study involves several critical ethical considerations, outlined comprehensively below:

- **Data Privacy and Protection:** Ensuring the protection of sensitive biodiversity data, particularly precise geographic locations of crayfish populations, is paramount to preventing exploitation or harm to vulnerable species. Detailed location data should be accessible exclusively to verified, registered users with validated credentials. Unregistered users should receive obscured spatial representations (such as hexagonal grids) to prevent misuse or unintended harm. The data management and sharing practices will adhere strictly to established international biodiversity conservation standards, data privacy regulations such as GDPR, and applicable national laws.
- **Data Quality, Accuracy, and Reliability:** The integrity of biodiversity databases is crucial for effective decision-making. WoC™ prioritizes expert-curated datasets to minimize inaccuracies that may arise from citizen science contributions. Clear and rigorous guidelines, along with structured validation procedures, are implemented to ensure that any citizen-generated data meets scientific quality standards. All submissions will undergo strict validation processes managed by qualified experts, thereby ensuring reliability and accuracy, reducing potential biases or errors, and maintaining scientific credibility.

- **Transparency and Open Science Principles:** WoC™ commits to fostering transparency by meticulously documenting all data sources, validation methods, analytical processes, and AI-driven modelling techniques. This approach promotes reproducibility and trustworthiness of the outcomes. WoC™ aligns with open science principles, ensuring responsible and equitable access to validated datasets for scientific research, education, and policymaking, balanced with data protection for sensitive species.
- **Compliance with Regulatory Frameworks:** The case study explicitly acknowledges and adheres to existing regulatory frameworks governing biodiversity management, including local, national, and international conventions and directives (e.g., Convention on Biological Diversity, EU Biodiversity Strategy, national environmental protection legislation). Adherence to these regulations ensures legal compliance and ethical responsibility in ecological research and biodiversity management.
- **Interdisciplinary Ethical Standards and Collaboration:** Given the interdisciplinary nature of the WoC™ project, integrating biology, GIS, statistics, and information technology, maintaining ethical standards across diverse disciplines is essential. Clearly defined ethical responsibilities, transparent communication, accountability, and integrity must be upheld within collaborative teams. Promoting mutual understanding of ethical guidelines across disciplines ensures smooth interaction, minimizes potential conflicts, and maximizes effective collaboration for ecological outcomes.
- **Responsible Use of Analytical and Predictive Models:** Analytical tools and predictive AI models used for invasive species tracking and ecological forecasting must clearly articulate their methodologies, assumptions, and potential limitations. Ethical AI usage requires transparency in algorithmic decision-making processes, data interpretation accuracy, and clear communication to stakeholders, avoiding misrepresentation or misuse of ecological information. The platform will emphasize responsible usage guidelines, including educating users on appropriate interpretation and application of analytical outputs.

By clearly defining and addressing these ethical considerations, WoC™ ensures robust, trustworthy, and ethically sound management and dissemination of biodiversity data.

2.2.7 Need for MultiGIS AI

The integration of MultiGIS AI into the WoC™ platform addresses significant limitations in current practices related to biodiversity data management and invasive species monitoring. The following rationale outlines why incorporating AI-enhanced MultiGIS capabilities is crucial:

- **Limitations of Current Platform Capabilities:** At present, WoC™ provides limited functionalities, predominantly restricted to static geographical visualizations. It lacks advanced analytical tools capable of extracting meaningful ecological insights or predictive modelling. Current analysis is manually intensive, and the absence of automation restricts scalability, responsiveness, and accuracy.

  The manual or semi-manual analysis is time-consuming and prone to human error, particularly when handling the extensive volume (over 106,000 records) and complexity of biodiversity data.

- **Need for Predictive Capabilities:** Invasive crayfish species pose significant ecological threats due to their aggressive nature, rapid reproduction, and adverse impacts on native biodiversity. Predictive modelling is essential to anticipate invasive species' spread, enabling timely and effective ecological interventions.

MultiGIS integrated with AI-driven modelling can systematically incorporate habitat characteristics, species-specific ecological traits, environmental data, and historical records to provide robust and accurate forecasts on invasive species distribution and future spread.

- **Enhanced Data Integration and Analysis:** AI-driven MultiGIS significantly enhances data integration capabilities. It allows for the seamless combination of geospatial habitat layers with existing biodiversity records to perform comprehensive spatial analyses and ecological assessments.

  Such integration supports advanced queries, real-time statistical analyses (e.g., min/max distributions, mean values, standard deviations), and customizable visualization outputs crucial for informed decision-making by stakeholders.

- **Real-Time Decision Support:** The dynamic nature of invasive species spread requires rapid analysis and dissemination of findings. AI-powered analytics can rapidly process and synthesize real-time data streams, enabling swift identification and response strategies by researchers and policymakers.

  Automated reporting functionalities and real-time analytical visualizations significantly improve response times and decision-making accuracy, essential for conservation and biodiversity protection efforts.

- **Improvement of Data Quality and Coverage Analysis:** MultiGIS AI supports advanced data quality checks, anomaly detection, and validation procedures, thereby improving the overall accuracy of the biodiversity database.

  Implementation of innovative tools like **cheCkOVER** (coverage analysis tool) relies on sophisticated GIS and statistical methods to evaluate how effectively current or newly uploaded datasets cover existing species distribution ranges. AI enhances these analyses by efficiently quantifying overlaps, identifying data gaps, and providing robust metrics for ecological studies.

- **Enhanced User Experience and Accessibility:** The integration of AI into the MultiGIS platform significantly improves user experience by providing a more intuitive, responsive, and intelligent interface. Advanced algorithms facilitate easier and more efficient interaction with complex ecological and geospatial data, enhancing platform adoption among scientists, policymakers, NGOs, and educational institutions.

  Additionally, incorporating offline capabilities enabled by AI-driven synchronization ensures accessibility for field researchers in remote or connectivity-limited regions, significantly improving data collection and analytical continuity in diverse operational contexts.

In summary, MultiGIS AI provides the WoC™ platform with transformative capabilities to overcome current data management, analysis, and prediction limitations, ensuring precise, reliable, and scientifically rigorous management of invasive crayfish data. The resulting platform substantially enhances the effectiveness and responsiveness of ecological monitoring and biodiversity conservation efforts.

2.2.8 Use cases

ID	WoC_UC_01
Title	Predictive Modelling of Invasive Crayfish Spread
Description	This use case involves leveraging AI-driven analytics integrated into the MultiGIS infrastructure to predict the potential spread of invasive crayfish species in Western Romania. The goal is to proactively manage biodiversity threats by anticipating invasions, allowing timely intervention and effective ecological management.
Addressed Business Needs	BN21
Addressed Desired Enhancements	FE1, FE2, FE3, FE4, FE5
Primary Actor	Ecologists, Environmental Authorities, Conservation NGOs
Preconditions	• WoC™ database populated with validated crayfish distribution data. • Geospatial data layers describing aquatic habitats and environmental characteristics available. • AI predictive modelling tools implemented and integrated within MultiGIS. • User is authenticated and has necessary permissions.
Postconditions	• Generation of predictive maps indicating potential areas of crayfish invasion. • Availability of detailed statistical analyses on spread patterns. • Comprehensive reporting (tables, graphs, and geographic visualizations) provided to authorized users.
Main Success Scenario	1. User logs into the WoC™ platform and selects "Predictive Spread Model." 2. User specifies parameters (species, geographic region, timeframe, environmental variables). 3. The WoC™ system queries its integrated MultiGIS database and extracts relevant ecological and geospatial data. 4. AI algorithms perform predictive analyses based on selected parameters, historical records, habitat suitability, and ecological variables. 5. The system generates a detailed predictive map and accompanying statistical summary.

	6. User reviews the output and optionally exports data for use in ecological management planning, policy formulation, or conservation strategy implementation.
Extensions	- **Insufficient Data Availability:** System alerts the user when chosen parameters result in inadequate data, prompting alternative parameter selection. - **Data Validation Failure:** If anomalies or potential inaccuracies are detected during analysis, the system notifies administrators to perform data validation. - **System Error or Failure:** In the event of technical failures during modelling, users receive clear notifications, and system logs the issue for technical review.
Frequency of Use	Monthly, or as required based on research or ecological management needs.
Status	Planned
Owner	UVT
Priority	High

ID	WoC_UC_02
Title	Analyzing Species Distribution Coverage (cheCkOVER)
Description	The cheCkOVER tool enables users to evaluate how comprehensively existing or newly provided datasets cover the known geographic range of crayfish species. This analytical capability is essential for ecological research validation, biodiversity monitoring, and reporting the significance and coverage of field studies.
Addressed Business Needs	BN22
Addressed Desired Enhancements	FE2, FE3, FE4, FE5

Preconditions	• WoC™ database contains validated species range and distribution data. • The user has a secondary dataset available, either pre-existing within WoC™ or newly uploaded by the user. • User is authenticated and has necessary permissions.
Postconditions	• Generation of detailed coverage reports, including maps and statistical tables. • Identification of overlap and coverage gaps, presented as percentage coverage and detailed geographic visualizations.
Main Success Scenario	1. User logs into the WoC™ platform and selects the cheCkOVER tool. 2. User specifies target species for range coverage analysis. 3. User selects secondary dataset (existing in WoC™ or newly uploaded). 4. The WoC™ platform queries its database to retrieve the primary species distribution range data. 5. cheCkOVER calculates overlap by comparing the selected datasets. 6. System outputs visual maps highlighting coverage overlaps and gaps, accompanied by a detailed statistical report. 7. User accesses, reviews, and optionally exports the generated data for research or reporting purposes.
Extensions	• **Dataset Error or Anomaly:** ○ The user is notified of any inconsistencies or potential inaccuracies, prompting manual verification by WoC™ experts. • **Incomplete or Incompatible Input:** ○ If user-uploaded datasets are incomplete or incompatible, the user receives clear guidance to correct or reformat input data. • **Technical Failure:** ○ In case of technical or processing errors, users receive immediate notifications, and the event is logged for internal review and correction.
Frequency of Use	Regularly, especially following new data uploads or significant database updates.
Status	Planned
Owner	UVT
Priority	High

2.2.9 Functional Requirements

Data Requirements

The WoC™ platform supports collecting, analyzing, and visualizing data on invasive crayfish species in Western Romania. It organizes this data into five main categories:

- **Source Info:** DOI, URL, citation.
- **Geolocation:** Latitude, longitude (WGS84), coordinate accuracy.
- **Crayfish Info:** Scientific name, locality status, year, extinction status, genetic data (e.g., GenBank codes).
- **Pathogen Info:** Detected pathogens/symbionts, methods, genotypes, eDNA analysis.
- **Additional:** Contributor name, confidentiality level.

Key Data Entities

Entity	Description
Species	Crayfish species being tracked
Observation	Individual or group sightings
Location	Geographical data for each observation
Pathogen	Disease/symbiont data linked to observations
Source	Origin of the observation (e.g., publication)
Contributor	Submitter of the data

Data Requirements Summary

- **DR1–DR3:** Store species details, geolocation, and observation dates for analysis and mapping.
- **DR4–DR5:** Record sources and contributor info with confidentiality controls.
- **DR6–DR7:** Include pathogen and genetic data for deeper analysis.
- **DR8:** Mark data as verified or pending expert review.

Functional Process Requirements

Functional process requirements

The WoC™ platform supports several interconnected functional processes designed to facilitate the collection, validation, analysis, visualization, and reporting of invasive crayfish data. These processes ensure that data are efficiently handled across their lifecycle, aligning with both user needs and scientific standards.

Context

The key functional processes are centered around enabling authorized users (scientists, conservationists, policymakers) to interact with the platform through data entry, analysis tools, and visual interfaces. These functions operate within a secure, user-authenticated environment.

Main Functional Processes

FRP1: Observation Data Ingestion

- **Input:** User-submitted forms containing species data (scientific name, date, geolocation, observer, genetic info).
- **Process Logic:** Validates required fields, checks for data consistency and format accuracy, flags for expert review if unverified.
- **Output:** Stores observation with metadata (source, contributor, validation status) into the platform's database.
- **Stored Data Accessed:** Species, Observation, Location, Source, Contributor.

FRP2: Geospatial Visualization and Mapping

- **Input:** Filter parameters (species, date range, validation status, location).
- **Process Logic:** Queries spatial database to render species distributions over a base map; for unregistered users, data is aggregated into hexagonal grids to protect sensitive coordinates.
- **Output:** Dynamic maps with overlays for habitat, risk zones, and historical spread patterns.
- **Stored Data Accessed:** Observation, Location, Species.

FRP3: Predictive Modeling (AI-driven)

- **Input:** Selected environmental layers and species data.
- **Process Logic:** Extracts relevant data, runs AI algorithms to forecast species spread based on historical and environmental variables.
- **Output:** Risk maps and trend analysis graphs presented via the user dashboard.
- **Stored Data Accessed:** Species, Observation, Location, Environmental Data.

FRP4: Overlap and Coverage Analysis (cheCkOVER)

- **Input:** One or more selected species datasets.
- **Process Logic:** Computes spatial and statistical overlaps between species distributions; identifies areas of data redundancy or absence.
- **Output:** Coverage maps, overlap percentages, and downloadable reports.
- **Stored Data Accessed:** Observation, Species, Location.

FRP5: User and Access Management

- **Input:** Registration details, user role assignment, login credentials.
- **Process Logic:** Validates credentials, assigns access levels (e.g., researcher, admin), logs user actions for auditing.
- **Output:** Authenticated sessions, access to tools and sensitive data based on roles.
- **Stored Data Accessed:** Contributor, Confidentiality Level.

FRP6: Reporting and Exporting

- **Input:** User request for data report (timeframe, location, species).
- **Process Logic:** Extracts relevant records, formats into tabular and visual outputs (e.g., charts, graphs, PDFs).
- **Output:** Downloadable datasets and analytics reports.
- **Stored Data Accessed:** Observation, Species, Source, Pathogen, Genetic Data.

Decomposition and Workflow

Each major functional process is modular and interacts with others through defined APIs and backend services. For example, FRP1 (data ingestion) feeds into FRP2 and FRP3, while FRP4 and FRP6 depend on FRP2 outputs. These interdependencies are managed through structured process logic and data validation pipelines.

2. 3. Case study 3: Archaeoastronomical analysis of heritage sites, Al-Hayit and Khaibar, Saudi Arabia.

2.3.1 Description

The Kingdom's Vision 2030 strategy emphasizes **cultural heritage tourism** as a key pillar for economic diversification. By identifying, documenting, and interpreting heritage sites in regions like Khaybar, this case study supports the development of new tourism destinations. Businesses involved in tourism infrastructure, heritage site management, and cultural experiences can leverage this data to design informed and sustainable offerings. This **case study focuses** on pathways in the northwestern Arabian regions that lead to the oases of Khaybar and Al-Hait, northeast of Medina al Munawara. These routes are flanked by numerous **drystone funerary structures**, dating to the third millennium BCE (Kennedy et al. 2021). These structures, classified as "keyhole" or "pendant" tombs (Figure 2.3.1) based on their distinctive shapes, are typically located on elevated terrain near the pathways (Dalton et al. 2022). So far, the studies carried out at the Khaybar Harrat monuments state that most of these 'keyhole' and 'pendant' tombs are perpendicular to the pathways (Kennedy et al. 2015), but **no systematic analysis on the topic has been performed**. Their presence emphasizes the Arabian Peninsula's role in ancient migration and pilgrimage when the region was more habitable and richer in vegetation.

Figure 2.3.1. Example of heritage sites in Saudi Arabia. ©Google Earth.

Just in the Khaybar oasis, approximately 9500 funerary structures have been identified, from which over 2800 are pendants (Kennedy et al. 2021). Given the vast number of these tombs, conventional identification methods—whether conducted on-site or via high-resolution satellite imagery—are time-consuming. In this sense, the proposed case study advocates for the application of AI techniques to accelerate the detection process and help with the extraction of key data, particularly orientation and horizon profiles, which are fundamental to archaeoastronomy (Ruggles 2015). The main **problem** is to identify the funerary structures and their orientations through AI and afterwards, explore possible patterns perhaps related to local topography or astronomical targets.

This approach would serve to conduct an analysis without the need to travel or face such demanding terrain as the deserts of the Arabian Peninsula (although on-site measurements will always be welcome). Nonetheless, these data ought to be corroborated by archaeological, topographic, and eventually, ethnographic evidence. This represents a new perspective to approach archaeoastronomy, applied to specific cases of great historical and cultural interest.

2.3.2 Existing infrastructure and current practice

Saudi Arabia has made significant strides in recent years in building a robust archaeological infrastructure and advancing its research practices, particularly in regions like Al-Ula and Khaybar. The **Royal Commission for Al-Ula (RCU)** has become a central hub for archaeological and cultural heritage research in the Kingdom. It oversees a strategic program that includes large-scale surveys, excavations, and multidisciplinary studies across the country. Their work incorporates state-of-the-art technologies, including remote sensing, 3D modelling, and GIS. In parallel, the **NEOM** project is pioneering the use of machine learning and AI-driven tools for archaeological surveys in northwestern Saudi Arabia. Their teams are using platforms like ArcGIS Pro and image recognition models (e.g., Mask R-CNN) to detect analyze large datasets. However, besides the fact that the developed tools are not focused on archaeoastronomy, **none of them are open source**. While the analysis can be done by external GIS tools (e.g., QGIS) and AI algorithms there is **no integrated automated solution** for large scale archaeoastronomical analysis.

2.3.3 Business needs

BN31. Automate detection and orientation measurement of funerary structures using AI and MultiGIS for archaeoastronomy

This business needs provides scalable tools for identifying and analyzing sensitive archaeological zones, as well as their relevance, helping **heritage preservation**.

- **Input Data:**
    - Multispectral Satellite images.
    - Research studies.
- **Processing & Integration:**
    - Identifying objects in satellite images, extracting them and analyzing each one w.r.t. to its orientation and other properties (e.g., size) by using QGIS.
- **Output & User Interaction:**
    - Users will be able to select in QGIS an area of interest and perform analysis.
    - Users will be able to visualize statistical plots and summary reports exportable in ???format.
- **Technical Requirements:**
    - Integration of Python scripts for AI and image processing in QGIS.
    - Implementation of user-friendly QGIS plugin.

By integrating AI and MultiGIS within QGIS this business need supports the development of digital heritage platforms and even virtual reconstructions. These can be shared through partnerships with museums, educational institutions, or tech companies offering immersive experiences such as AR/VR tours relying on accurate digital reconstruction based on the provided analysis.

2.3.4 Engaged users

First and foremost, the **academic and research communities** stand to benefit. Archaeologists will find the proposed methodologies useful for identifying and analyzing large-scale spatial data. Archaeoastronomers, who explore the relationship between ancient structures and celestial phenomena, will be especially interested in the orientation measuring tools and analyses.

Saudi universities and research institutions—such as King Saud University, King Abdulaziz University, and the Saudi Commission for Tourism and National Heritage (SCTH)—might be key users of this research. These institutions have shown increasing interest in digital archaeology and the integration of GIS and remote sensing into heritage studies. The proposal's focus on the Khaybar and Al-Hait regions aligns with ongoing archaeological efforts in the Kingdom.

Governmental bodies, especially the **Ministry of Culture and the Heritage Commission**, would find the research relevant for heritage management and preservation. The identification and documentation of thousands of funerary structures through machine learning could significantly enhance the national archaeological registry and inform conservation strategies. Moreover, the ability to analyze these sites remotely is particularly valuable in a country where many heritage locations are in remote or environmentally challenging areas. This aligns with Saudi Arabia's Vision 2030 goals, which emphasize cultural heritage as a pillar of national identity and a driver of tourism and economic diversification.

Cultural heritage organizations and museums within Saudi Arabia, such as the **National Museum in Riyadh or the Al-Ula Royal Commission**, may also engage with the research. The insights into ancient funerary structures and their potential astronomical alignments could enrich museum exhibits, offering visitors another interpretation and a more immersive and scientifically grounded narrative of the Kingdom's ancient past. These findings could also support the development of heritage tourism routes, particularly in regions like Khaybar, which are being promoted as part of Saudi Arabia's growing cultural tourism sector.

Finally, **local communities** in and around Khaybar and Al-Hait might play a big role in the research. These communities may have oral histories, cultural practices, or traditional knowledge that can complement the scientific data, offering a more holistic understanding of the sites. Engaging local populations not only enriches the research but also fosters a sense of ownership and pride in their heritage, which is crucial for sustainable preservation efforts.

2.3.5 Ethical requirements

We must ensure that sensitive data—such as the exact locations of burial sites—is not publicly disclosed in ways that could lead to looting or unauthorized access. Although the dataset is publicly accessible via Google Earth, we should refrain from disclosing the precise locations of each structure. Instead, we should limit the information provided to archaeoastronomical data – such as orientation –linked to the identification numbers assigned to each detected object. Ideally, data should be stored and shared in collaboration with Saudi institutions, ensuring local control and long-term stewardship.

Even though the research emphasizes remote sensing, any future fieldwork must be conducted with **minimal environmental impact**. The desert ecosystems around Khaybar and Al-Hait are fragile, and archaeological sites can be easily damaged by unregulated access. Ethical protocols should be in place to ensure that any on-site investigations are non-invasive and approved by the relevant heritage authorities.

2.3.6 Need for MultiGIS AI

The Khaybar and Al-Hait regions contain **thousands of funerary structures**, including keyhole and pendant-shaped tombs, scattered across vast and often inaccessible desert terrain. Traditional GIS methods rely heavily on manual digitization and visual interpretation of satellite imagery, which is **extremely time-consuming and prone to human error** when applied at this scale. AI can automate the detection and classification of these features, dramatically increasing efficiency and consistency.

Determining the orientation of thousands of structures manually (in-situ or through tools such as Google Earth Pro) is not only labor-intensive but also introduces variability in measurement. AI can be trained to **automatically extract orientation vectors** from satellite imagery, ensuring uniformity and enabling large-scale statistical analysis.

The harsh desert environment of northwestern Saudi Arabia makes fieldwork logistically challenging and expensive. AI-driven remote sensing allows researchers to **conduct high-resolution analysis without physical presence**, reducing the need for extensive field campaigns while still producing reliable data. This is particularly valuable for preliminary surveys or in politically or environmentally sensitive areas.

In the Khaybar region of Saudi Arabia, where thousands of prehistoric funerary structures are dispersed across arid and rocky terrain, the use of **multi-wavelength remote sensing** can significantly enhance archaeological detection and analysis. Among the most valuable spectral ranges, infrared and thermal imaging could help reveal buried or eroded features by detecting subtle differences in soil, moisture, and heat retention. **Radar** could penetrate dry sand and highlight surface roughness, while **hyperspectral** data distinguishes between construction materials and geological features. Integrating these wavelengths into a MultiGIS platform allows for a layered, multidimensional analysis of Al-Hait and Khaybar's funerary landscape, and thus enhancing AI algorithms with richer input data, improving detection accuracy.

2.3.7 Use cases

ID:	Archaeoastronomy_UC_01
Title:	Automate detection and orientation measurement of funerary structures using AI and MultiGIS for archaeoastronomy.
Description:	Automated detection, geometric analysis, and cultural interpretation of prehistoric funerary structures in the Khaybar region of Saudi Arabia. It integrates AI-based image recognition, automated orientation measurement, and statistical archaeoastronomical analysis within a MultiGIS platform.
Primary Actor:	AI-based image recognition system

Preconditions:	• Access to high-resolution multispectral satellite imagery. • A labelled training dataset of known structures with their locations. • AI model capable of detecting relevant features. • GIS environment prepared for data integration (e.g., QGIS, ArcGIS Pro). • Defined area of interest (AOI) in the regions.
Postconditions:	• A geospatial dataset of detected tombs with orientation attributes. • Visualizations of orientation vectors and spatial distributions. • Exportable data for further statistical analysis of the orientations.
Business needs:	• Manual detection is time-consuming and infeasible at scale; automation enables rapid processing of large areas. • Reduces human error and subjectivity in identifying features. • Supports expansion to other regions using the same or similar model architecture. • Provides foundational data for cultural heritage, i.e. site protection, tourism planning, and academic research.
Main Success Scenario:	• The AI model is deployed on satellite images of the Al-Hait and Khaybar regions. • It scans the images and identifies and classifies features matching the geometric and spectral characteristics of keyhole and pendant tombs. • Detected features are converted into vector polygons and georeferenced. Image enhancing when poor resolution. • The orientation analysis module calculates the azimuth of each tomb's main axis and stores it in the GIS database. • Horizon profiles are modeled using DEMs to assess visibility • The results are reviewed, validated, and integrated into the MultiGIS platform. • Results of azimuth and declination are visualized through diagrams, curvigrams, and maps, and interpreted in light of cultural and historical context. • The system generates a report summarizing detection statistics, spatial distribution, and confidence levels.
Extensions:	• Natural rock formations or modern structures may be misclassified as tombs. • Eroded or partially buried tombs may not be detected. • Performance may degrade in new regions or under different lighting/terrain conditions. • Cloud cover or low-resolution imagery may limit detection accuracy. • A validation interface may be needed for expert review and correction.

Frequency of Use:	Depending on survey scope and data availability.
Status:	In development
Owner:	UVT
Priority:	HIGH

2.3.8 Functional requirements

Data requirements

Six main data categories are identified.

Key Data Entities

Entity	Definition
Structure	Represents a funerary structure detected in satellite imagery. In this case study, it can be circular shape, pendant or keyhole.
SatelliteImage	The multispectral satellite image used for detecting structure
OrientationData	The orientation measurements (azimuths) of detected structures, used for archaeoastronomical analysis.
HorizonData	The height of the horizon or horizon profile of detected structures, used for archaeoastronomical analysis.
DEM	Represents Digital Elevation Models used for terrain analysis
Context	Represents the region, historical period, etc.

Attribute Definitions

Structure

- StructureID: Unique identifier for the structure.
- Shape: Geometric shape of the structure (e.g., keyhole, pendant).
- Location: Geospatial coordinates of the structure.

- Orientation: Azimuth orientation of the structure.
- DetectionConfidence: Confidence level of the structure detection.

SatelliteImage

- ImageID: Unique identifier for the satellite image.
- Resolution: Resolution of the satellite image.
- DateCaptured: Date when the image was captured.
- Source: Source of the satellite image.

OrientationData

- OrientationID: Unique identifier for the orientation data.
- StructureID: Identifier for the associated structure.
- Azimuth: Orientation angle of the structure relative to true north.

DEM

- DEMID: Unique identifier for the DEM.
- Region: Region covered by the DEM.
- Resolution: Resolution of the DEM.
- Source: Source of the DEM.

Context

- ContextID: Unique identifier for the cultural context.
- Region: Region associated with the structure.
- HistoricalPeriod: Historical period of the structure.
- DOI: available research related

Data Requirements Summary

- **DR1:** Obtain satellite image and enhance its resolution if necessary.
- **DR2:** Obtain a list of cropped images with the identified objects.
- **DR3:** Obtain data on the analysis of each identified objects.
- **DR4:** Obtain an aggregated report of all identified objects.

Functional Process Requirements

Key Functional Processes:

- **FRP1: Data Entry** – Users select an area of interest in the GIS platform.
- **FRP2: Object identification**– User select an AI algorithm for object identification and execute it through the interface. AI identifies the objects.
- **FRP3: Object Analysis** – Each object is analyzed using image processing and AI.
- **FRP4: Reporting** – Aggregated statistical reports are generated for all analyzed objects.

References

Chelmsford Garden Community Consortium (2023). ‘Development Framework Document – The best of toem & country, the best of neighbourhoods & nature, A shared Place’. chelmsford-garden-community-development-framework-document-january-2023.pdf

Domisch, S. et al. (2015). *Near-global freshwater-specific environmental variables for biodiversity analyses in 1 km resolution*, Scientific Data. https://doi.org/10.1038/sdata.2015.73

Ion, M. et al. (2024). *World of Crayfish™: a web platform towards real-time global mapping of freshwater crayfish and their pathogens*, Bioinformatics and Genomics. https://doi.org/10.7717/peerj.18229

Kennedy, M. et al. (2021). *Dating the pendant burials of north-west Arabia: First radiometric results from the Khaybar Oasis, Saudi Arabia,* Arabian Archaeology and Epigraphy. https://doi.org/10.1111/aae.12199

Rodríguez-Antón, A. et al. (2018). *Estimating the Reliability of Digital Data Acquisition in Cultural Astronomy: The Case of Roman North Africa*, Journal of Skyscape Archaeology, https://doi.org/10.1558/jsa.34685

Ruggles, C.L.N. (2015). *Analyzing Orientations*. In: Ruggles, C. (eds) Handbook of Archaeoastronomy and Ethnoastronomy. https://doi.org/10.1007/978-1-4614-6141-8_26