

This draft document is currently under revision by the European Commission (EC) and has not yet been validated or approved by the EC. The content provided herein is subject to change, and the information presented may not represent the final position or official stance of the EC.



**EIS**  
Exploration  
Information  
System

# D 3.1: Design Report

EIS Toolkit & QGIS EIS Wizard

Version 1.0

**Lead Beneficiary: GISPO**  
04 / 2023

**Niko Aarnio<sup>1</sup>, Timo Aarnio<sup>1</sup>, Miikka Kallio<sup>1</sup>, Johanna Torppa<sup>2</sup>, Bijal Chudasama<sup>2</sup>, Sakari Hautala<sup>2</sup>, Juho Laitala, Ina Storch<sup>3</sup>, Peggy Hielscher<sup>3</sup>, Michael Steffen<sup>3</sup>, Andreas Knobloch<sup>3</sup>, Paavo Nevalainen<sup>4</sup>, Jukka Heikkonen<sup>4</sup>, Fahimeh Farahnakian<sup>4</sup>, Maija Lespinasse<sup>4</sup>, Martiya Sadeghi<sup>5</sup>, Ildiko Antal<sup>5</sup>, Guillaume Bertrand<sup>6</sup>, Blandine Gourcerol<sup>6</sup>**

<sup>1</sup>GISPO

<sup>2</sup>GTK

<sup>3</sup>Beak Consultants

<sup>4</sup>UTU

<sup>5</sup>SGU

<sup>6</sup>BRGM



Funded by  
the European Union

## Disclaimer

The content of this report reflects only the author's view. The European Commission is not responsible for any use that may be made of the information it contains.

*under revision by the European Commission*

## Document information

<b>Grant Agreement / Proposal ID</b>	101057357
<b>Project Title</b>	Exploration Information System
<b>Project Acronym</b>	EIS
<b>Scientific Coordinator</b>	Vesa Nykänen (vesa.nykanen@gtk.fi) – GTK
<b>Project starting date (duration)</b>	1 May 2022 (36 months)
<b>Related Work Package</b>	WP 3
<b>Related Task(s)</b>	Task 3.1
<b>Lead Organisation</b>	GISPO
<b>Contributing Partner(s)</b>	GTK, BEAK, UTU, SGU, BRGM
<b>Due Date</b>	30.04.2023
<b>Submission Date</b>	28.04.2023
<b>Dissemination level</b>	PU

## History

Date	Version	Submitted by	Reviewed by	Comments
07.07.2022	0.1	Andreas Knobloch		Template
18.04.2023	0.2	Timo Aarnio		Draft
26.04.2023	1.0	Andreas Knobloch		Revision

## Table of contents

1.	Introduction .....	9
1.1	General remarks .....	9
1.2	EIS WP3 and Task 3.1.....	9
1.3	EIS WP2 and Task 2.1.....	9
1.4	Review of EIS Methodology.....	10
1.4.1	Mineral Prospectivity analysis.....	10
1.4.2	Mineral System Approach .....	10
1.4.3	From Methodology to Software Tools .....	10
1.5	Related Existing Software .....	11
1.5.1	Software Tools with User Interface.....	11
1.5.1.1	QGIS.....	11
1.5.1.2	advangeo 2D Prediction .....	11
1.5.2	Python Libraries.....	12
1.5.2.1	Pandas .....	12
1.5.2.2	Matplotlib.....	12
1.5.2.3	Plotly .....	12
1.5.2.4	Scikit-learn.....	12
1.5.2.5	Gdal .....	12
1.5.2.6	GeoPandas .....	13
1.5.2.7	Tensorflow .....	13
1.5.2.8	Keras.....	13
1.5.2.9	Statsmodels.....	13
1.5.2.10	Rasterio .....	13
1.5.2.11	Qt / PyQt .....	14
2.	Software Requirements .....	15
3.	Software Concept.....	16
4.	Software Design .....	18
4.1	EIS Toolkit .....	18
4.1.1	Raster processing .....	18
4.1.2	Vector processing.....	19
4.1.3	Data transformations .....	20
4.1.4	Exploratory Data Analysis.....	20
4.1.5	Prediction .....	22
4.1.6	Validation .....	23
4.1.7	Utilities .....	24



- 4.2 EIS QGIS Plugin..... 24
  - 4.2.1 EIS QGIS Processing Algorithms ..... 25
  - 4.2.2 EIS QGIS Wizard ..... 26
  - 4.2.3 Plugin – Toolkit –interface design ..... 27
- 5. Software licensing and development process ..... 29
  - 5.1 Software licensing..... 29
  - 5.2 Development process ..... 29

*Under revision by the European Commission*

## List of figures

Figure 1: Mineral predictive mapping (MPM) flowchart for EIS.....	17
Figure 2: Design template for EIS tool in the EIS QGIS wizard.....	25
Figure 3: Usage of EIS tools in the model builder in QGIS .....	26
Figure 4: EIS QGIS Plugin – EIS Toolkit architecture .....	28

*Under revision by the European Commission*

# List of tables

Table 1: Raster processing tools of initial EIS toolkit .....18

Table 2: Vector processing tools of initial EIS toolkit .....19

Table 3: Data transformation tools of initial EIS toolkit .....20

Table 4: Exploratory data analysis tools of initial EIS toolkit .....21

Table 5: Prediction tools of initial EIS toolkit .....22

Table 6: Validation tools of initial EIS toolkit .....23

Table 7: Utilities of initial EIS toolkit.....24

*under revision by the European Commission*



## Abbreviations and Acronyms

Acronym	Description
WP	Work Package
GIS	Geographic Information System
EIS	Exploration Information System
GUI	Graphical User Interface

## Summary

The Deliverable D3.1 provides a description of the contents and general structure of the 1) EIS Toolkit and 2) EIS QGIS Wizard implemented in this project.

## Keywords

Software Design, Mineral Prospectivity Modelling, Mineral Predictive Mapping, QGIS, Toolkit, Artificial Intelligence, Toolkit, Wizard

# 1. Introduction

## 1.1 General remarks

This document will give an overview of the background, concept and design of the EIS Toolkit and the QGIS EIS Plugin developed in WP3 of the EIS project. EIS Toolkit is a Python library of functions customized for prospectivity analysis. The QGIS EIS Plugin consists of two parts: The EIS Wizard and the EIS Processing Toolbox. The EIS Wizard is a user interface guiding through different steps of prospectivity analysis and providing the possibility to organize data and track the processing workflow. The EIS Processing Toolbox is the QGIS interface to the EIS Toolkit functions, and it can be used outside the EIS Wizard for building custom prospectivity analysis workflows in QGIS, for instance, with the QGIS model builder.

Additional algorithms were originally planned to be included in this document, but are described in Deliverable 3.2, instead.

## 1.2 EIS WP3 and Task 3.1

The main objective of WP3 is the development of a GIS (Geographical Information System)-based Exploration Information System (EIS) for predictive mapping of mineral resources. EIS does not have a strict definition but can be characterized as an environment for performing data analysis and modelling, for managing data and other information, and for representing results in various forms.

In Task 3.1, the general structure of the "EIS QGIS Wizard" will be developed, and main contents of the "EIS Toolkit" and the "EIS Processing Toolbox" are defined and described in this document. The planning of the general structure of "EIS QGIS Wizard" involves both infrastructure and content related questions. A successful wizard-like solution requires creating an organized structure between each workflow step and identifying the possible needs user may have during the workflow. Our goal was to identify and design the components to be included in the wizard and generate a list of tools and functionalities that will be included in it. Only tools with clear roles/uses are included in the wizard design. We also reviewed existing software and explored the possibilities to utilize the tools / plugins already existing in QGIS and elsewhere as open-source code.

The idea of the "EIS Toolkit" is to act as a comprehensive Python library of independent functions relevant for mineral prospectivity analysis. These tasks include mainly predictive mapping and data integration via mathematical modelling, but also some general data processing and analysis. Also, tools for evaluation of the goodness of the models and modelling results will be included for efficient decision making in the identification and prioritization of exploration targets. "EIS Toolkit" will contain implementations of existing and new algorithms. Emphasis is given to exploring the applicability of modern machine learning methods, such as convolutional neural networks, the use of which in mineral prospectivity modelling is still at its infancy.

The structure of and interfaces to the functions included in the library have been planned so that the functions can be smoothly integrated to other software, such as the "EIS QGIS Wizard". Existing Python libraries have been reviewed and are used to minimize the coding effort. Only functions that are not directly available in other libraries included in the toolkit are implemented.

## 1.3 EIS WP2 and Task 2.1

In WP2, three types of mineral systems are studied to define proxies that can be expressed as mappable variables and used in prospectivity analysis as explanatory variables. In the "EIS QGIS Wizard", we integrate these proxies as a guidance for the user, what data to use for the mineral prospectivity analysis and how to pre-process it,

regarding to the mineral system the user has chosen in the Wizard. The tree mineral systems and all of the mapable proxies are described in the report of Task 2.1.

## 1.4 Review of EIS Methodology

### 1.4.1 Mineral Prospectivity analysis

Predictive modelling of areas favorable for the existence of undiscovered mineral deposits is usually called *mineral predictive modelling*, *mineral prospectivity modelling* or *mineral potential modelling*. Instead of modelling, some researchers use the term *mapping*. A more general term to describe this complex multi-stage process is *mineral prospectivity analysis*. The number of different terms already suggests that there is no universally recognized best way or set of established tools for producing information about where the deposits most likely are.

Mineral systems of any type are connected with specific characteristic geological features, often distinct from the surrounding geology. As these features have certain physical and chemical properties, mineral systems often can be distinguished from the surroundings by integrating spatial geochemical and geophysical information from different sources (vegetation, till or stream sediment samples, airborne or ground geophysical measurements, even satellite-based measurements). Mineral systems are complex, however, and the surrounding geology varies in different places, which makes it difficult to generate universally applicable models for even a certain type of mineral system. The approaches to carry out modelling (integrating data from different sources) include unsupervised and supervised modelling methods. Unsupervised methods are used, when there are not enough training data, i.e., known deposits, in the study area. These approaches require using a priori expert knowledge, i.e., qualitative estimates, on how different geochemical and geophysical quantities are related to the deposit and its surroundings. Supervised methods make use of training data to generate a mathematical model for relating measurement data to the existence of deposits. Both approaches require a lengthy procedure of data processing prior to actual modelling to transform the original measurements into a form that best distinguishes the mineralisation. In machine learning nomenclature this procedure is called *feature engineering*.

### 1.4.2 Mineral System Approach

There are certain geological phenomena related to the formation of each type of mineral system, that leave a footprint on the deposit location and its surroundings. The mineral system approach in prospectivity analysis means that the geological processes of the entire mineral system over time, instead of just the deposit itself, is carefully considered when selecting and processing the input data in prospectivity analysis. An EIS contains, in addition to prospectivity modelling tools, a library of mineral system components and the related observable features for each type of mineral system to facilitate selection and processing of input data.

### 1.4.3 From Methodology to Software Tools

When developing software for performing mineral prospectivity analysis, we must consider the different use cases, techniques and workflows as well as different types of input data.

Use cases include, for instance

- Governmental organization working on a regional scale to determine the overall mineral potential
- Exploration company working on intermediate scale to determine sites for detailed exploration activities

- Research institute working on all scales to determine the geology of the study area and test different computational techniques for prospectivity analysis

There are a multitude of techniques for all the stages of mineral prospectivity analysis. Users have their own favourite techniques and workflows for reaching the final prospectivity model. Thus, all different techniques must be available as individual functions for the user to be able to combine them according to their needs.

To make it as easy for the user as possibly, different types of input data must be considered. Data can be given as a raster or vector file, or as a numeric table. Gridded data can have varying resolution and pixel alignment. Coordinate systems vary between the datasets. Data type can be a string, float or integer. Pre-processing the data to be consistent and compatible for mineral prospectivity modelling can be tedious, and proper tools reduce the workload of this obligatory step.

## 1.5 Related Existing Software

### 1.5.1 Software Tools with User Interface

#### 1.5.1.1 QGIS

QGIS ([www.qgis.org](http://www.qgis.org)) is a free and open-source Geographic Information System (GIS) software available for all major operating systems (Linux, macOS, Windows). It supports a wide variety of geospatial data formats and APIs and provides tools for visualizing, creating, editing, analysing and publishing data.

The functionality of QGIS can be extended with plugins that can be installed directly from the user interface of QGIS either from a public registry of plugins or from a zip file. The QGIS user and developer community is very large and development, training and other services related to QGIS are available from many companies.

The QGIS EIS Plugin and EIS Processing Toolbox are implemented for QGIS.

#### 1.5.1.2 advangeo 2D Prediction

The advangeo<sup>®</sup> 2D Prediction software is a modelling and prediction software developed by Beak Consultants GmbH for analysing and modelling spatial data with artificial intelligence within the Esri ArcGIS environment. This software is highly flexible and can be used for multiple purposes. Almost any natural event or phenomenon based on spatial data can be predicted using advangeo<sup>®</sup>, including erosional phenomena, mineral prediction, or prediction of coal fires.

The software provides a complete workflow for designing and processing parametrized models. All working steps, such as problem definition, data selection, data processing, modelling parameters, and results, are easily comprehensible and can be saved. Additionally, integrated data and the results of modelling and prediction can be easily visualized in ESRI ArcMap. In the modelling step, the user can choose between different methods, such as Multi-Layer Perceptron (MLP), Self-Organizing Maps (SOM), Weights of Evidence (WofE), Fuzzy Logic, and Random Forest (Regression and Classification).

The advangeo<sup>®</sup> software offers several extensions that facilitate the import and processing of geospatial data. For instance, one extension enables the import of grids of geophysical data and the calculation of derivatives to determine slope, aspect, and/or curvature. The geophysical data is then scaled and resampled to match the resolution of a predefined base raster, producing an input layer suitable for predictive modelling. Another extension is designed for remote sensing data, allowing the import of multispectral or hyperspectral data. For example, Sentinel-2 data from a Copernicus download can be automatically stacked, clipped, and resampled for use in mineral mapping, utilizing USGS spectral libraries or custom spectral libraries from ASD/PSR+. A third extension, the Tectonic Extension, allows for the import of Esri shape files with faults, with the ability to calculate

fields for length and direction, classify the faults, and generate grids containing faults with a specific length or direction, along with their Euclidean distance. Additional extensions are available for erosion modelling, geological maps, geochemical data, and processing DEM derivatives.

## 1.5.2 Python Libraries

As the language of implementation is Python, existing stable Python libraries are used to avoid unnecessary coding. These libraries are general purpose libraries for machine learning, statistical computations, GIS data processing, and visualization.

### 1.5.2.1 Pandas

Pandas is a widely used open-source Python library for manipulating and analyzing data. It offers an efficient way to store and manipulate data through powerful data structures. Pandas offer a wide range of functions for cleaning, wrangling and reshaping data. It is capable of reading and writing data from various sources, such as CSV files, Excel spreadsheets and SQL databases.

### 1.5.2.2 Matplotlib

Matplotlib is a comprehensive and most widely used visualization library for Python. It is fast and can export visualizations to all popular image formats. The library provides the capability for creating visualizations in a variety of styles, such as line graphs, histograms, scatter plots and pie charts. Additionally, 3D charting is made possible with Matplotlib.

### 1.5.2.3 Plotly

Plotly is another graphing library for Python. One of the main advantages of Plotly is that it makes creating interactive visualizations possible and easy. Compared to matplotlib, Plotly's interface is more high-level, and it offers an extensive number of predefined plot types.

### 1.5.2.4 Scikit-learn

Scikit-learn, also known as sklearn, is a popular open-source machine learning library for Python. It is built on NumPy, SciPy and matplotlib, making Scikit-learn fast and scalable for large datasets. The library provides efficient tools for predictive data analysis and is designed with an easy-to-use API that makes experimentation with different algorithms and models easy.

### 1.5.2.5 Gdal

GDAL (Geospatial Data Abstraction Library) is a widely used open-source library for working with different kinds of geospatial data formats. It supports both raster and vector formats and besides supporting conversion between formats it can be used for data translation and transformation as well.

GDAL is a de facto component in most of the modern geospatial software and libraries such as QGIS, ArcGIS and Google Earth Engine.

#### 1.5.2.6 GeoPandas

GeoPandas is a Python library that extends the functionality of Pandas by providing support for geospatial data. GeoPandas allows the use of spatial operations, such as geocoding, spatial joins and analysis. It is built on top of widely used Python libraries (Shapely, Fiona, PyProj) that allows the usage of various spatial data formats, geometric objects and coordinate transformations. GeoPandas is commonly used for data visualization, geospatial analysis, spatial data integration and geocoding.

#### 1.5.2.7 Tensorflow

Tensorflow is a machine learning library building and training machine learning models. It supports various types of models, including convolutional-, deep- and recurrent neural networks. Tensorflow allows training of large models on distributed computing systems, due to its high scalability and efficiency. It uses a computational graph to represent mathematical operations and data flow, allowing it to compute gradients and optimize parameters during training efficiently. In addition, Tensorflow comes with tools for data preprocessing, visualization, and model deployment.

#### 1.5.2.8 Keras

Keras is a high-level neural networks API designed to run on top of lower-level deep learning frameworks, such as Tensorflow. It provides a user-friendly interface for building and training deep neural networks. Keras allows its users to define and configure complex neural network architectures with ease by using simple and consistent API. Additionally, it comes with a range of utilities and functions for data preprocessing, regularization, and model evaluation. Keras provides a way to quickly experiment and prototype with different neural network architectures and hyperparameters, without tweaking low-level details of the underlying deep learning framework.

#### 1.5.2.9 Statsmodels

Statsmodels is a Python package that is primarily used for statistical analysis, modeling, and visualization. It is an open-source library that contains a variety of statistical models, including linear regression, time series analysis, generalized linear models, and more. Statsmodels also provides tools for exploring and visualizing data, such as summary statistics, hypothesis testing, and graphics.

#### 1.5.2.10 Rasterio

Rasterio is a Python library for working with geospatial raster data. It is built on top of GDAL and aims to provide a more user-friendly and pythonic yet as powerful alternative for processing raster data. All basic operations such as raster reading, writing, reprojecting, resampling, masking, windowing, and spatial indexing are available in the library. Rasterio can be used both as a Python library in scripting and developing other packages as well as directly in the command line, using Rasterio's command line interface called "rio".

### 1.5.2.11 Qt / PyQt

QT is a framework for creating graphical user interfaces (GUIs). It is cross-platform capable and supports all the major operating systems, such as Windows, macOS, Linux, Android and iOS. QT provides libraries and tools for developing GUIs for both desktop and mobile applications. QT also comes with a visual design tool, QT Designer, which provides the ability to create user interfaces with an easy-to-use drag-and-drop interface.

PyQT enables the usage of QT functionality directly from Python code, by providing Python bindings for the QT framework. PyQT provides easier and more pythonic way to write code for the QT framework.

under revision by the European Commission

## 2. Software Requirements

In the first months of the project Gispo conducted a series of stakeholder interviews to increase understanding about the work related to MPM. Altogether five people, each from a different partner organization, were interviewed. The partners included in the interviews were Beak, SGU, BRGM, GTK and Malayer University.

The interviews had constructed of prepared questions about input data, tools, and workflow. In addition to the questions a bit of time for more open discussion about the topic was reserved for each participant.

The results of the interviews were compiled into a document that was distributed to the project partners and used as a base for requirements for the software.

Under revision by the European Commission



### 3. Software Concept

Implementation notes document on how to integrate toolkit into QGIS

The tools in the EIS Toolkit will be integrated into the EIS QGIS Wizard, a guided end-to-end implementation with a visual user interface for performing mineral prospectivity analysis. “EIS QGIS Wizard” will be running under the open GIS platform, QGIS, thus bringing the developed methods and tools openly accessible to everyone. In addition to the tools in the “EIS Toolkit”, the wizard will include general data exploration, processing, and visualization tools. The functionality of “EIS QGIS Wizard” can roughly be divided as follows:

- **Data pre-processing tools**, which will enable transforming primary geoscientific data to synthetic features representing mineralization processes and/or mineralization-related/favourable geological features.
- **Data analysis tools**, which will enable the visualization and exploratory data analysis of the primary and pre-processed data, as well as testing of geological hypotheses connected with the primary and synthetic features.
- **Data integration and modelling tools**, which include statistical, analytical and machine learning based data analysis tools for integrating information from different types of input data. The tools can be used, depending on the case, to either generate mineral prospectivity models or to extract from the datasets relevant features related to the mineral deposits. Both supervised and unsupervised machine learning techniques will be included.
- **Model testing tools**, which are used for testing the prediction capability and/or the success rate of the model. Several tools are included for different use cases.
- **Tools for creating visual outputs**, which produce maps representing data (input data, refined data, derived data), plots describing obtained results, and models obtained from data analysis/integration. Tools for visualizing data and results in spatial frame are important because both the data and the examined phenomenon (i.e., mineral deposit occurrence) are spatial in nature.

The “EIS QGIS Wizard” will be implemented so that additional algorithms or tools can be added later to each section. This allows this project to act as an opening statement for building a complete GIS-based data analysis pipeline for implementation of machine learning methods on geospatial datasets for mineral prospectivity modelling. Because the source codes of the “EIS QGIS Wizard” as well as the “EIS Toolkit” will be distributed openly in GitHub or in a similar software distribution platform, contributions from numerous developers globally are allowed, making the development, maintenance and updating of such a complex system possible.

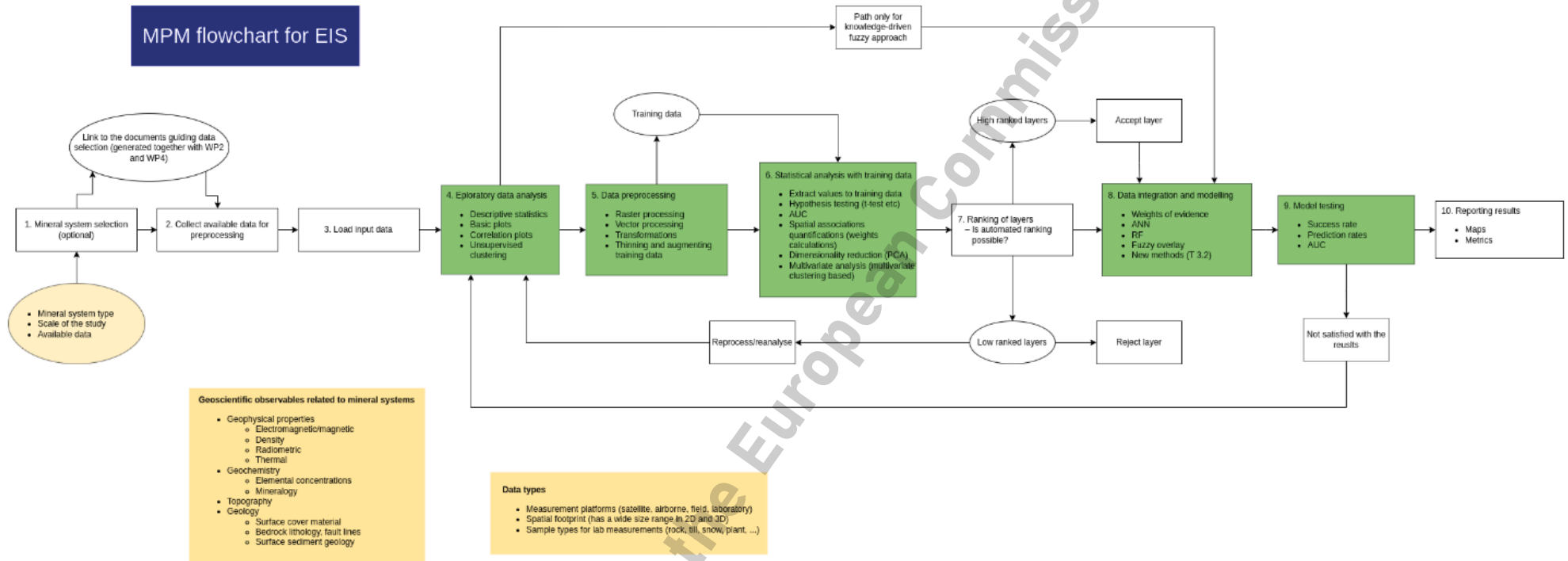


Figure 1: Mineral predictive mapping (MPM) flowchart for EIS

under revision by the European Commission

## 4. Software Design

### 4.1 EIS Toolkit

The algorithms of EIS Toolkit will be divided into different modules based on their role in mineral prospectivity mapping workflow. These modules group the tools to make the structure of the toolkit intuitive for users.

#### 4.1.1 Raster processing

Raster processing tools consists of a set of commonly used processing tools applicable for raster datasets. These tools are used to pre-process, and prepare data for exploration and modelling tools.

**Table 1: Raster processing tools of initial EIS toolkit**

Tool name	Tool description	Tool implementation status
Reproject raster	Reprojects a raster to another coordinate reference system.	Ready
Resample raster	Resamples a raster to a new resolution and pixel size.	Ready
Clip raster	Clips a raster with a vector dataset.	Ready
Snap raster	Snaps/aligns raster grid with a grid defined by another raster.	Ready
Window raster	Extracts a window from a raster.	Ready
Unify rasters	Unifies (resamples, reprojects, aligns and optionally clips) a list of rasters to correspond a base raster.	Ready
Reclassify raster	Reclassifies the pixel values of a raster.	In progress
Create constant raster	Creates a raster with given constant pixel value.	In progress
Extract values from raster	Extracts pixel or attribute values from a specified location(s).	In progress
Unique conditions grid	Creates a new raster from a set of rasters. The pixel values of	In progress

	the new raster correspond to unique combinations when the input rasters were stacked.	
Aspect	Creates a new raster where pixel values show the slope direction.	In progress
Slope	Creates a new raster where pixel values show the slope steepness (first derivative).	In progress
Curvature	Creates a new raster where pixel values show the slope curvature (second derivative).	In progress

## 4.1.2 Vector processing

Vector processing module consists of a set of commonly used processing tools applicable for vector datasets. These tools are used to pre-process and prepare data for exploration and modelling tools.

**Table 2: Vector processing tools of initial EIS toolkit**

Tool name	Tool description	Tool implementation status
Reproject vector	Reprojects a vector to another coordinate reference system.	Ready
Interpolate vector	Interpolates a vector using selected interpolation method.	In progress
Extract shared lines between two polygons	Identifies and selects the shared boundary lines of adjacent polygons.	Not started
CoDA – log ratio transforms	Transform compositional data into Euclidean space.	Not started
Rasterize	Converts a vector into a raster.	In progress
Density computations	Calculates spatial density of a feature within an area.	In progress

Distance computations	Calculates the distance between two or more features.	In progress
Calculate geometries	Generates new geometries or attributes based on the spatial characteristics of existing geometries.	Not started

### 4.1.3 Data transformations

Data transformations module consists of data transformation manipulation tools used in pre-processing. These tools are commonly used before the data can be used in in modelling tools.

**Table 3: Data transformation tools of initial EIS toolkit**

Tool name	Tool description	Tool implementation status
Binarize	Replaces values with 0 and 1 based on specified threshold.	In review
Winsorize	Replaces extreme values to fit within specified range.	In review
Logarithmic	Transforms input data with log2 or log10.	In review
Sigmoid	Transforms input data to values between 0 and 1 based on the sigmoid function.	In review
Linear	Transform input data based on mean and standard deviation. Able to also transform based on specified minimum and maximum values.	In review

### 4.1.4 Exploratory Data Analysis

The exploratory data analysis module provides various tools for exploring, analysing, and visualising data.

**Table 4: Exploratory data analysis tools of initial EIS toolkit**

Tool name	Tool description	Tool implementation status
PCA	Computes the principal components for the input data.	Ready
Auto encoders	Transforms/encodes the unlabelled input data.	Not started
K-means clustering	Clusters data points into k clusters based on selected attribute.	Not started
Descriptive statistics	Calculates and summarizes common statistics for the input data (Mean, quantile, standard deviation, relative standard deviation and skewness).	In progress
Plot scatterplot	Generates a scatterplot for visualizing the relationship between two continuous variables.	In progress
Plot histogram	Generates a histogram plot for visualizing the distribution of a continuous variable.	In progress
Plot correlation matrix	Generates a matrix plot for visualizing the pairwise correlations between a set of continuous variables.	In progress
Plot PCA	Generates plots for visualizing the results of PCA.	Ready
Plot boxplot	Generates a plot for visualizing the distribution of a set of continuous data.	In progress
Plot heatmap	Generates a plot for visualizing the distribution of a variable or set of variables across different categories.	Not started

Plot parallel coordinates	Generates a plot with polylines and vertices to visualize high-dimensional data.	Not started
Plot density scatterplot	Generates a special scatterplot and bins data to visualize data with large number of points.	Not started

## 4.1.5 Prediction

The prediction module contains models that can predict or classify the input data. The usage of these functions usually requires a level of preprocessing or exploratory data analysis as a preliminary step.

**Table 5: Prediction tools of initial EIS toolkit**

Tool name	Tool description	Tool implementation status
Fuzzy overlay	Combines datasets using fuzzy logic and the selected overlay method (and, or, product, sum, gamma).	In review
Weights of evidence	A data-driven modelling method that calculates the relationship between presence and absence of a feature and predictor variables.	In review
Artificial neural networks (ANN)	A machine learning model that identifies complex patterns and relationships within input data.	In progress
Convolutional neural networks (CNN)	A deep learning model that identifies spatial patterns and structures in raster data.	In progress
Logistic regression	Predicts the probability of a binary outcome based on predictor variables.	Not started
Random forests and decision trees	Random forests construct multiple decision trees to control overfitting and improve prediction accuracy.	In progress

Multilayer perceptron	A feedforward ANN for continuous and categorical data prediction.	In review
-----------------------	---	-----------

## 4.1.6 Validation

The tools in validation module are used for evaluating the performance of models. These tools are essential to understand the accuracy and performance of the models and identify how to get the best possible predictions.

**Table 6: Validation tools of initial EIS toolkit**

Tool name	Tool description	Tool implementation status
Area under curve (AUC)	Measures the performance of a classification by calculating the area under ROC curve.	Ready
Plot ROC curve	Measures the performance of a classification visually. Plots true positive rate against false positive rate for different thresholds.	Ready
Plot success rate curve	Visualizes the success rate (true positive rate).	Ready
Confusion matrix	Presents true positives, false positives, true negatives and false negatives in a tabular form.	Not started
RA plot	Visually shows if a model's performance is better than a random guess.	In progress
Conditional independence test	Statistically measure the independency and relevance of predictive factors used in a model.	Not started
Validate ANN	Validates the performance of an ANN.	Not started
Uncertainty estimator	Estimates the uncertainty of model predictions.	Not started



## 4.1.7 Utilities

In addition to pre-processing, exploratory data analysis and modelling tools, the EIS toolkit also contains miscellaneous utility functions. These functions are used to perform basic operations, such as reading and writing data, handling nodata and converting data formats.

**Table 7: Utilities of initial EIS toolkit**

Tool name	Tool description	Tool implementation status
Read raster	Read and open a raster dataset.	In progress
Read vector	Read and open a vector dataset.	In progress
Write raster	Write data to a raster dataset.	In progress
Write vector	Write data to a vector dataset.	In progress
Check raster grids	Checks that the set of input rasters have matching CRS, pixel size, pixel alignment and optionally bounds.	Ready
Set nodata	Rewrite nodata value in raster metadata.	In progress
Replace nodata	Replaces old nodata values with a new value.	In progress
Read CSV to geodataframe	Converts csv file into geodataframe.	Ready
Convert raster band to dataframe	Converts a band of a raster into a dataframe.	Ready

## 4.2 EIS QGIS Plugin

EIS QGIS Plugin is the graphical user interface (GUI) designed for EIS Toolkit. Because the GUI is created as a QGIS plugin, users won't need any additional software to visualize their data and can leverage many useful features of QGIS.

EIS QGIS Plugin is comprised of two main components: EIS QGIS Wizard and EIS QGIS Processing Algorithms. Initially, EIS QGIS Wizard was supposed to be the only GUI for EIS Toolkit. However, it was later decided that implementing individual tools from the toolkit as QGIS Processing Algorithms would be beneficial to allow users to access the tools without using EIS QGIS Wizard and to leverage the Model Designer feature in QGIS.

An important note to make is that EIS QGIS Plugin does not perform any computations or execute EIS algorithms. All processing will be handled in EIS Toolkit and EIS QGIS Plugin will merely order computations and fetch results.

## 4.2.1 EIS QGIS Processing Algorithms

EIS QGIS Processing Algorithms are implemented using the processing algorithm framework of QGIS. This means that the algorithm interfaces are programmed using a specific template and structure and that the only design work needed is to label and order input parameter fields for each algorithm.



Figure 2: Design template for EIS tool in the EIS QGIS wizard

All QGIS processing algorithms, including EIS QGIS Processing Algorithms, will be available in QGIS Model Builder, an existing graphical modelling interface in QGIS. The Model Designer enables users to create custom models by connecting various processing algorithms available in QGIS with data inputs and outputs. This feature provides a flexible and visual way for users to take advantage of EIS QGIS Processing Algorithms.

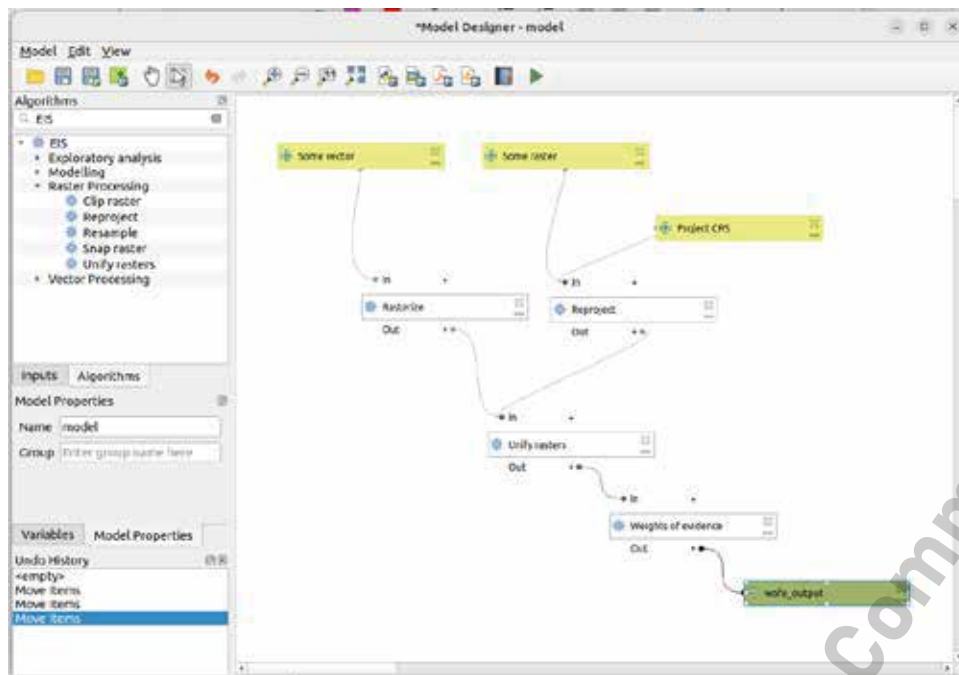


Figure 3: Usage of EIS tools in the model builder in QGIS

## 4.2.2 EIS QGIS Wizard

EIS QGIS Wizard is the primary GUI component of EIS QGIS Plugin. EIS QGIS Wizard will facilitate project management, group related tools and steps and guide more inexperienced users through an MPM workflow.

The clearest and most important way to understand EIS QGIS Wizard is through UI windows. The wizard is planned to have the following windows:

- **Preprocessing window**, which facilitates processing mappable proxies of geological processes from primary/raw geological data.
- **Exploration window**, which shows different methods to explore the selected data. Visualizations and analysis reports will be embedded in the exploration window.
- **Modeling window**, which allows users to select a modeling method of their choice and do final preparations for their model data. This window will likely also involve validation, unless it is decided validation needs its own window.
- **Settings window**, which allows users to set and configure miscellaneous settings, such as EIS Toolkit installation location.

Project management is facilitated by utilizing the tree-like, hierarchical structure of layer groups in QGIS. Groups will be automatically created and populated when a project is started and when processing algorithms are run. For simplicity and because it was seen sufficient, an EIS project is coupled with a QGIS project file. All settings, layers and project state are saved in a QGIS project file.

The UI of EIS QGIS Wizard will be implemented using PyQt5 Python library, both programmatically and using Qt Designer, a software made specifically to design Qt graphical user interfaces. Most of the design work so far has been carried out in UI design workshops involving programmers and stakeholders, and the initial implementation has been created based on comments by an UI designer, geologists and PyQt programmers.

### 4.2.3 Plugin – Toolkit –interface design

As EIS Toolkit is separate from EIS QGIS Plugin, there is need to define an interface between the plugin and the toolkit. The interface will be created between EIS QGIS Processing Algorithms and EIS Toolkit to have a clearly defined and general interface. EIS QGIS Wizard will execute EIS QGIS Processing Algorithms and connect to the toolkit in that way.

The interface implementation sets two requirements for the user: Firstly, the user needs to have both EIS QGIS Plugin and EIS Toolkit installed, and secondly the user needs to set EIS Toolkit installation location for the plugin when using the plugin for the first time.

under revision by the European Commission

### EIS QGIS Plugin - EIS Toolkit architecture

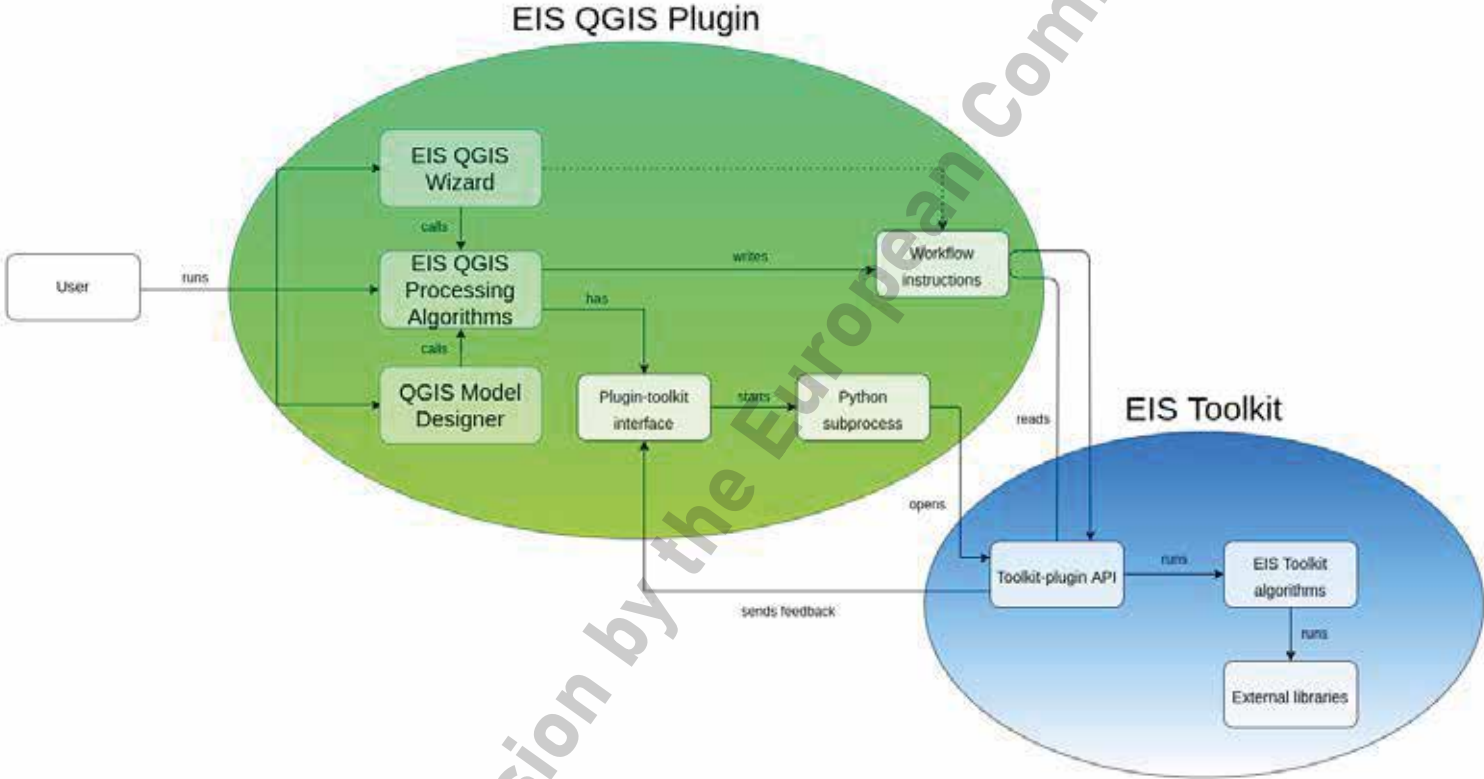


Figure 4: EIS QGIS Plugin – EIS Toolkit architecture

## 5. Software licensing and development process

### 5.1 Software licensing

The licence chosen for EIS Toolkit and QGIS EIS Wizard is EUPL 1.2 which is a so called copyleft licence. A copyleft license is a type of open-source software license that ensures that the software and its derivatives remain free and open source.

The used licence does not restrict the use of EIS Toolkit as a part of a commercial software if the licence conditions of EUPL are complied with. In practice this means that any modifications to the EIS Toolkit must be published with the same licence.

### 5.2 Development process

As multiple organisations take part in the development work for the EIS Toolkit and the QGIS EIS Wizard, a software development process taking this into account had to be agreed upon and documented. With the help of other project partners, Gispo defined the contributing guidelines and documented them into GitHub, which is used to host both the EIS Toolkit Python library and EIS QGIS Plugin. The guidelines instruct with setting up a development environment, how to write functions/code, how to test and finally how to submit new functionality as Pull Requests. After a Pull Request has been submitted it will be reviewed by one or more of the developers of the project.

The two repositories can be found from:

- [https://github.com/GispoCoding/eis\\_toolkit/](https://github.com/GispoCoding/eis_toolkit/)
- [https://github.com/GispoCoding/eis\\_qgis\\_plugin](https://github.com/GispoCoding/eis_qgis_plugin)

The EIS Toolkit repository has been opened for public access and the EIS QGIS Plugin repository will be made public at a later stage of development.

## Conclusion

This document describes the design phase of the EIS Toolkit and the EIS QGIS Plugin. We investigated the possibilities of current software used for similar tasks as well as Python libraries that could be used in our own implementation. A list of required functionalities was compiled and prioritized for implementation. We interviewed different types of users of different software available today and carefully studied the workflow they are using. The design work continued with a series of workshops arranged by GTK, Gispo and Beak focusing on the user interface and user experience of the EIS QGIS Plugin.

A process for the development of EIS-related software was also designed and tested as part of the task. The process is documented and can be found from the GitHub repository of EIS Toolkit.

The results of the project will be published with a free and open-source licence. The licence chosen for EIS Toolkit and EIS QGIS Plugin is EUPL 1.2, which is a GPL-compatible copyleft licence.

The work continues with implementing the designed software, testing, and reacting to the results of the testing.

under revision by the European Commission

## Bibliography

Nykänen, V., Lahti, I., Niiranen, T., & Korhonen, K. (2015). Receiver operating characteristics (ROC) as validation tool for prospectivity models — A magmatic Ni–Cu case study from the Central Lapland Greenstone Belt, Northern Finland. *Ore Geology Reviews*, 71, 853-860.  
doi:10.1016/j.oregeorev.2014.09.007

under revision by the European Commission