Short Acronym and full Title	**Artificial Intelligence (AI) / machine learning for analysis of site characterisation subsurface data**		
Type of activity	☒R&D	☐Strategic Study	Knowledge Management – covered by a separate committee and template
Budget estimation (total budget in M€, i.e ~ 1.5 M€)	**€3M**	Duration of the WP (in months)	36
Links with EURAD SRA / Roadmap Themes (if multiple choices, indicate the primary link in bold – maximum 3)	☐Programme Management (Theme 1) ☐Pre-disposal (Theme 2) ☐ Engineered Barrier Systems (Theme 3) ☒**Geoscience (Theme 4)** ☐ Disposal facility design and optimisation (Theme 5) ☒ Siting and Licensing (Theme 6) ☒ Safety Case (Theme 7)		
Links with EURAD SRA topics (if multiple choices, indicate the primary link in bold – maximum 3)	The techniques utilised/developed within the proposed work package support a variety of objectives within Theme 4 of the EURAD Roadmap, including: - **4.1.1 Develop a model of the host rock and surrounding geological environment, including distributions of rock types, geometry and properties of structural features, geotechnical properties and the hydrogeological and hydrogeochemical environment (Site descriptive model).** - 4.1.2 Describe bedrock transport properties (aqueous and gas transport, advection/dispersion, diffusion) including retention (sorption, matrix diffusion) of different geological material. 4.3.2 Assess the nature of future climate change and landscape evolution and its potential impacts on THMC conditions in the repository host rock (including the repository) and surrounding formations (Climate change)		
SRA drivers (maximum 3)	☐Implementation Safety	☐Tailored Solutions	☒Scientific Insight
	☒Innovation for Optimisation	☐Societal Engagement	☒Knowledge Management
Objective (What) – 1 sentence	To develop and evaluate AI and machine learning methodologies for integrating subsurface datasets, with an emphasis primarily on deep geological boreholes, to enhance geological understanding and predictive modelling in support of Geological Disposal Facility (GDF) siting.		

Justification: impact / innovation / added-value (Why) – bullet points or short paragraph **(maximum quarter of a page)**	Safety cases, engineering designs, programmes and work packages to support the siting of Geological Disposal Facilities generate and require vast quantities of high-quality, verified and validated subsurface data and information. To achieve this, Waste Management Organizations (WMOs) subsurface characterisation programmes from early characterisation through to concept design stage are underpinned by a broad range of data acquisition campaigns. The data collected in these campaigns are highly varied, with variable sampling, resolution parameterisation and quality amongst a range of factors. Subsurface data may fall into quantitative measurements (e.g. a point density measurement, core mineralogical analysis), relative measurement (e.g. Seismic reflection data) or qualitative (e.g. a core photograph, outcrops). The requirement to integrate disparate and esoteric data types is a common and longstanding challenge in geosciences. This workflow often relies on qualitative or model-based analysis and interpretations presented in long form reporting to produce models and conclusions. AI and machine learning methods offer powerful tools for integrating diverse and complex subsurface datasets. These models excel at convolving different data types, for example images, text and numbers, by converting them into data space and then learning relationships between them. For example, such techniques have been shown to link digitised core photographs with descriptive metadata, or wireline log data, enabling the generation of one from the other. This approach supports both forward and reverse predictions, allowing for the synthesis or interpretation of missing or incomplete data based on learned patterns, together with an efficient check of the consistency of the different data response. The databases aggregated in the siting process for a GDF are in many cases globally unique due to the extensive quality control of all incoming data and the predominant emphasis on sampling a spatially and stratigraphically discrete host rock unit and strict control on 3D positioning of samples lends these datasets ideally towards AI convolutional models. The results of successfully integrating all subsurface data into single object can provide an enhanced understanding of the behaviour of, and ultimately the characterisation of GDF host rock units. Moreover, integrating all available data and being able to conduct mathematical operations allows to extract better insight and more value from the existing data by identifying underlying parameter correlations essential for the building of models and digital twins
List of planned tasks / subtasks with % of effort per task (5% increments) **(Maximum 10 bullets)**	**Task 1 (5%): Project management** **Task 2 (10%): Knowledge Management** **Task 3 (25%) Develop a comprehensive set of best practice, processes, standards and structure for borehole data including:** Focussing principally on an archetypal deep characterisation borehole, this task focuses on integrating all typically available multimodal data related to site characterisation (geological, geophysical and geothechnical) in a unified framework and structure. • Based upon data supplied by a participating WMO, identify, aggregate and analyse an archetypical subsurface deep borehole geological datasets. If multiple datasets from several organisations are received, the most suitable input data will be defined. Some datasets will be retained for future "blind-test" reference.

- Perform a complete inventory of all available data (e.g. gamma, FMI, resistivity, core photos, porewater data, reports etc) Develop a database structure for the storage of borehole data within a database

- referencing all data collected from the archetypal site characterisation datasets.

- Identify how all data (including ascii, image files, records, engineering parameters etc) may be structured according to a standard reference frame e.g. depth

- Identify how data of variable sample interval, sample frequency, instrument resolution and integration volume may be consistently stored. Generate a reference database to underpin future analysis and evaluation.

- Validate the structure accommodates all identified data types.

- Incorporate mechanisms for data import/export in machine-readable formats (e.g., JSON, XML, HDF5).

Task 4 (35%) Identification and assessment of AI/ML techniques for multimodal subsurface data analysis and interpretation

This task aims at applying and further developing AI/ML techniques to support the data integration and analysis of data stemming from Task 3. Subsequently the objective is to extract geological, geochemical, geotechnical and structural insights and interpretations relevant to the geological disposal site characterization.

- Review of state-of-the art AI/ML methodologies to identify the most suitable techniques for data integration

- Adaption of methods or further development of existing methods or development of new methods as they apply to reference data. (e.g. multimodal transformer, convolutional, geostatistics, graph generative architectures)

- Development of preliminary data pipelines/ workflows to reference data including standardisation of units, scales, resolutions and formats.

- If data is available, identify methods of spatial integration of multiple boreholes from the same geological site.

- If seismic data is available, identify methods of correlation spatial correlation of boreholes using seismic data

- Application of AI/ML models for key data consistency tasks: prediction of missing measurements, detection of inconsistencies, pattern identification and cross-data correlation.

- Extract valuable information and address the explainability of the AI/ML methods.

Task 5 (25%) Workflow validation, demonstration and proof-of-concept

- Execute and demonstrate full lifecycle workflow on complete dataset as input for AI/ML workflows

- Demonstrate validation of used methods. Use AI/ML for geological assessment and multimodal characterisation. Use explainable AI tools such as SHAP and saliency maps. Assess prediction accuracy

	• Identify value added over traditional methods • Identify methods of quantifying and reporting uncertainty • Produce workflow and analytical methods map and report value adding cases (e.g. geological model construction) • Demonstrate workflows, cross-validate, concepts and methods and knowledge of pre-trained models on a "blind-test validation dataset". Identify limitations and provide recommendations. • Report and export pre-trained models for broad usage along EURAD community
List of expected outcomes linked to the identified SRA drivers **(Maximum 6 bullets)**	• Development of an AI ready multimodal subsurface data structure for the integration of logs, images, text descriptions, laboratory measurements, hydraulic tests and geochemical datasets. • A greater understanding of the requirements of a potential data layer or structure to underpin AI/ML analysis • Identification of the applicability of AI/ML models, workflows and techniques for the integration of subsurface data related to site characterisation for deep geological disposal. • A greater understanding of the methods of reporting and quantifying uncertainty, demonstration cases and pre-trained tools to be used across the EURAD community (WMOs and Scientific)
Deliverables **(Maximum 6 – including the prescribed deliverables)**	• Reference Borehole Data Repository for future R&D and analysis Reference Dataset Integration: rules and workflows for adapting datasets to the proposed structure and standardized subsurface and borehole Data Framework report • Outline AI/ML (including adapted models) Workflows for subsurface data analysis and interpretation along with documentation • Output and provision of pre-trained models for future adoption and development. Reusable code and models across EURAD-2 community to support geo-models and site characterization. Potential input for EURAD/WMO digital twin initiatives. Including benchmarking and testing report. Recommendations and limitations • Outcome and impacts report, Mmember Sstates and end users • Initial state of the art report (Mth 6), Final state of the Art Report (3 years)
Critical input requirements & identified risks	• Archetypical borehole dataset and ideally reference datasets for the "Blind Test" of derived methodologies (partial datasets have been already published) • Potential constraints due to data confidentiality or incomplete metadata
Major achievements expected by end of Year 2	• Complete Integration of Archetypal Dataset • Identification and plan for creation of a preliminary AI/ML Pipeline Operational • (Potential) Prototype to identify methods of spatial data integration • Early proof-of-concept demonstration

OFFICIAL

(Go/No Assessment)[1] **(Maximum 5 bullets)**	• Reusable codes and models across EURAD-2 community
(Optional - Explain what is out of the scope?)	Perform THMC simulations or safety assessment studies Development of organisation specific guidelines, standards or variations to the database structure Development of site-specific AI analytical workflows Waste repository design or engineering optimization Site specific geomodels
List of preliminary interested organisations as partners in the WP contributing effort; **% of effort (person months, by College)**	**Nuclear Waste Services (NWS**), Andra (WMO), ASNR (TSO), FTMC (TSO), GRS (TSO), NRG PALLAS (TSO), BGE (WMO), Covra (WMO), Nagra (WMO), RHK (WMO), SKB (WMO), Posiva (WMO), Sogin S.p.a (WMO), PSI (RE), NNL (RE), LEI (RE), SCK CEN (RE), SIIEG NASU (RE), UFZ (RE), VTT (TSO), CVUT (RE), DELFT (RE), Galson Sciences (RE), Geological Society of Finland (RE), GFZ Helmoltz Centre for Geosciences (RE), INGECID (RE), Lancaster University (RE), Atkins Realis (RE), UniPi (RE) ENRESA (WMO) as end-user
If applicable - links with previous projects / work packages	This WP builds on and complements prior EURAD-1 efforts in geoscience and modelling. It contributes directly to EURAD-2 aims on digitalization, reproducibility and usage of advanced tools for extracting value from vast datasets. WP is relevant to: EURAD-DONUT EURAD-2 HERMES EURAD-2 DITOCO2030
WP Preparation Team (1 member per College) contact (**organisation + person**, email)	**David Eastwell, Nuclear Waste Services (WMO).** David.eastwell@nuclearwasteservices.uk **Nikolaos Prasianakis, PSI, (RE).** nikolaos.prasianakis@psi.ch **Wilfried Pfingsten, SITEX.Network, (TSO)** wpfingsten@pop.agri.ch

---

[1] EC budget being only allocated for the first 2 years, each work package progress will be reviewed at the end of Year 2, to assess its continuation based on the total budget that EURAD-2 will be granted.