Annals of Nuclear Energy 196 (2024) 110212

Contents lists available at ScienceDirect

Annals of Nuclear Energy

journal homepage: www.elsevier.com/locate/anucene

Intelligent nuclear decommissioning solution: Code for site characterization and management of overall surveys

Hyungi Byun [a,b], Jong Dae Park [a], Sihyun An [c], Jinu Kim [c], Juyub Kim [a,c,*], Doo Yong Lee [a], Bongsoo Lee [b]

[a] FNC Technology Co., Ltd., 32 Fl., 13 Heungdeok 1-ro, Giheung-gu, Yongin-si, Gyeonggi-do 16954, Korea
[b] Department of Energy Systems Engineering, Chung-Ang University, 84, Heukseok-ro, Dongjak-gu, Seoul 06974, Korea
[c] Department of Industrial Engineering, Yonsei University, 50, Yonsei-ro, Seodaemun-gu, Seoul 03722, Korea

ARTICLE INFO

Keywords:
Nuclear decommissioning
Site characterization
Software development
Artificial intelligence
Data management

ABSTRACT

The site characterization should be conducted during the entire decommissioning period to verify the compliance with regulatory guidance and to demonstrate the safety for site release. A technical manual called Multi-Agency Radiation Survey and Site Investigation Manual (MARSSIM) has been widely used as a reference for the site characterization. Historically, it has required enormous resources because of extensive data handling. And there is a possibility of a human error because of the expert's judgement reliance. In this study, software called COde for Site characterization and Management of Overall Surveys (COSMOS) has been developed with Artificial Intelligence (AI) models. One of the AI model is possible to revolutionary reduce time for finding keyword related contamination, and the other manages database by digitalizing drawings. Also, other survey modules are possible to conduct related all of statistical tests. In the end, COSMOS would be utilized as a total management solution for site characterization.

1. Introduction

Each Nuclear Power Plants (NPPs) has an Operation License Period (OLP) and it must be decided whether the OLP should be extended or terminated before the end of the OLP. If it is decided to terminate, then there are three decommissioning strategies: immediate dismantling (DECON), deferred dismantling (SAFSTOR), and entombment (ENTOMB) (IAEA, 2005).

ENTOMB is the strategic planning for preserving NPP in perpetuity in a safe manner. The objective of DECON and SAFSOR is to ensure adherence to dismantling regulatory guidance and safety for unrestricted or restricted site release. Radiological surveys on the site are critical during the decommissioning process. Surveys called site characterization should be conducted in NPPs as well as other nuclear facilities. If site characterization confirms adherence to regulatory guidance for site release, then the site can be reused in greenfield or brownfield applications after decommissioning. Greenfield is the unrestricted site release that is back to the state prior to the construction after decommissioning, whereas brownfield is the restricted site release that is to reuse for plant construction after decommissioning.

International Atomic Energy Agency (IAEA) has set the regulatory guideline of site release below 0.3 mSv/y and conceded that optimized radiation protection for below 0.01 mSv/y may not be necessary (IAEA, 2006). In the United States, the Nuclear Regulatory Commission (NRC) has stipulated that radiation should not exceed 0.25 mSv/y on an unlimited site release for the total effective dose of the average determined group and minimize residual radioactivity to as low as reasonably achievable (ALARA) (NRC, 2021). In Europe, Germany has stipulated that any person should not be exposed to residual radioactivity beyond an effective dose of 0.01 mSv/y from (OECD/NEA, 2014). Spain has set regulatory guidelines that limited the exposure of the effective dose of 0.1 mSv/y. Buildings and structures should also satisfy these criteria (CSN, 2007). In Korea, the Nuclear Safety and Security Commission (NSSC) Notice No. 2021-11 set the limit of residual radioactivity to 0.1 mSv/y for both unrestricted and restricted site release (NSSC, 2021).[1]

Every country sets distinct regulatory guidelines. However, site

* Corresponding author at: FNC Technology Co., Ltd., 32 Fl., 13 Heungdeok 1-ro, Giheung-gu, Yongin-si, Gyeonggi-do 16954, Korea.
E-mail address: yubjoo@fnctech.com (J. Kim).

[1] For unrestricted release, a licensee should consider all possible exposure pathways by residual activities and the individual effective dose should not exceed 0.1 mSv per year. If it is expected that the unrestricted release criteria cannot be satisfied (restricted release), a licensee may reuse the site on a limited basis only if the individual dose does not exceed 0.1 mSv per year with restrictions to make the exposure by residual activities as low as reasonably achievable.

https://doi.org/10.1016/j.anucene.2023.110212
Received 19 July 2023; Received in revised form 26 October 2023; Accepted 29 October 2023
Available online 3 November 2023
0306-4549/

characterization is typically performed to demonstrate safety for the unrestricted or restricted site release. Site characterization during decommissioning of nuclear facilities generally follows Multi-Agency Radiation Survey and Site Investigation Manual (MARSSIM) developed by NRC, Environmental Protection Agency (EPA), Department of Energy, and Department of Defense in the United States. The procedure of site characterization consists of Historical Site Assessment (HSA), scoping survey, characterization survey, remedial action support survey, and Final Status Survey (FSS). The HSA is used to collect existing information concerning the site and its surroundings and is used to identify the potential sources of residual radioactive material and determine whether a site is contaminated or not. This provides details for scoping and characterization survey designs. The scoping survey is used to classify all or part of the site and evaluate whether the survey plan can be optimized for use in the characterization survey or FSS. The characterization survey is planned based on HSA and scoping survey results. This type of survey typically details the radiological environmental characterization of the area. If an area is adequately characterized and has concentrations of residual radioactive material above Derived Concentration Guideline Levels (DCGLs), a remediation plan should be prepared. The remedial action support survey is performed simultaneously when remediation is being conducted and guides the remediation in real time. Finally, the FSS was used to demonstrate compliance with release criteria. This survey was the major concern of MARSSIM (NRC, 2000).

HSA requires considerable time and manpower because of extensive data and documents. However, HSA is prone to human errors because many procedures are subjective. Because HSA rely on historical records, documents, and reports, which may be incomplete, inaccurate, or inconsistent. The quality and availability of these records can vary significantly, making it challenging to obtain a complete and reliable basis of the site's history. And interpreting historical data requires subjective judgment. Different individuals may interpret the same historical documents or artifacts differently, leading to variations in the assessment's conclusions. Other surveys are required to manage various data, such as measurement equipment, or device, survey result, and statistical test result. Thus, site characterization requires management in the evaluation process and results in the long term. In this study, COde for Site characterization and Management of Overall Surveys (COSMOS) was developed to reduce the burden of engineers using the Artificial Intelligence (A.I.) model to support Radiological Survey and Site Investigation (RSSI), and manage assessment. This study introduced COSMOS program and revealed that all surveys can be performed through a test case.

2. Existing software for the site characterization

Comparing existing software revealed that only a few software programs have been developed in this field. Therefore, this study introduced four representative software and their functions.

2.1. Visual sample plan (VSP)

VSP software was developed by the Pacific Northwest National Laboratory (PNNL) to support the development of a defensible sampling plan based on statistical sampling theory and the statistical analysis of sample results to support confident decision-making. VSP combines site, building, and sample location visualization capabilities with optimal sampling design and statistical analysis strategies and can be used to perform calculations to determine the number of locations in which measurements should be performed or where samples should be collected (PNNL, 2001). This software shows through User Interface (UI) the sampling point on the drawing for measurement of the soil and buildings subject to perform site characterization and provides it. However, there is no function to perform statistical test to evaluate the satisfaction of the site release based on the measurement results of the point.

2.2. COMPASS

Oak Ridge Associated Universities developed COMPASS software to help the user make informed decisions when designing final status radiological surveys. COMPASS simplified the application of the statistical tests by performing statistical calculations and providing prospective power curves to determine the level of confidence the user is willing to accept for a particular number of measurements or samples for a survey unit. After performing the FSS, COMPASS can be used to assess data and compared with release criteria (RSICC, 2002). This software has with function to calculate the number of sampling points, especially it has the advantage of being able to apply the requirement of MARSSIM by dividing the number of interested nuclide into one or above two. However, it does not deal with the distinction between contaminations or not for physical areas, and the survey unit generation cannot be performed on them.

2.3. Spatial analysis and decision assistance (SADA)

SADA was developed by university of Tennessee to measure the location that can be schematized on a photograph of the actual site or on a cadastral map. SADA presents detailed representation of the site size, number of surface layers, depth, grid shape and allows to determine design errors of preprocessed values (calculated by DCGL, etc.) presented by MARSSIM (Robert Stewart et al., 2009). SADA can show through UI the sampling point on the drawing for measurement of the soil and buildings subject to perform site characterization and provides it. However, there is no function to perform statistical test to evaluate the satisfaction of the site release based on the measurement results of the point.

2.4. ProUCL

U.S.EPA developed ProUCL software as the latest update of statistical software package for analysis of environmental data sets with and without non-detected observations. ProUCL is a comprehensive statistical software package with statistical methods and graphical tools to address many environmental sampling and statistical issues. ProUCL can be used to evaluate the probability of exceeding the upper bound of the gray region using an upper confidence limit (EPA, 2013). It has the advantage of being able to derive results from statistical test in various ways to prove the satisfaction of the site release criteria, which is the purpose of site characterization. However, there is a probability of human error because users should directly distinguish which areas are contaminated based on documents with evidence.

2.5. RESRAD

Argonne National Laboratory of United States developed RESRAD to estimate radiation doses and cancer risks to an individual on the site within the boundary of the primary contamination in soil, off site, or both on-site, and off site. RESRAD was used to extrapolate or interpolate the results of a single deterministic run to compute DCGLs. DCGLs are activity concentrations of each radionuclide in the contaminated medium that ensure that the predicted dose by each radionuclide and their progenies stay within the specified dose limit over the time horizon (NRC, 2020). This software was developed for the purpose of calculating dose based on various exposure scenarios, therefore it is optimized for the calculation of DCGL required by site characterization. However, it is impossible to view contaminated areas on drawing to users or show sampling points on the UI, and there is no function to perform statistical test.

Table 1 reveals that existing codes have limited functions such as schematization of the sample point on drawings, calculation of sample

Table 1
Existing software related to the site characterization.

Software	Organization	Functions
VSP	Pacific Northwest National Laboratory	✓ Schematization of sample point on drawings
COMPASS	Oak Ridge Associated Universities	✓ Calculation of sample point for statistical test Providing prospective power curves
SADA	University of Tennessee	✓ Schematization of sample point on drawings Allowance checking a design error (e.g. DCGL)
ProUCL	U.S. Environment Protection Agency	✓ Evaluate probability of exceeding the upper bound of the gray region
RESRAD	Argonne National Laboratory	✓ Calculation for residual radioactivity and derivation of DCGL regarding radioactive nuclides

point for the statistical test, and allowance checking a design error. Existing software cannot support site characterization as integrated software. Furthermore, database deployment for overall surveys is essential for managing long-term data. In the COSMOS, a lot of documents and drawings by the NPP to be evaluated can be entered to distinguish contaminated areas through AI, and based on this, sampling points can be calculated and shown to users through UI. In addition, if user input the measurement results for each interested nuclides according to the sampling point, the COSMOS automatically calculates whether the site release criteria, which is the purpose of site characterization, are satisfied and indicates the results. Therefore, we embedded all these functions into the COSMOS program. The structure and function of COSMOS are defined.

3. Methodology and algorithm

3.1. Software structure

When the decommissioning of nuclear facilities is finalized, then the first step of site characterization is HSA. HSA is performed to determine whether the facility and surrounding environment are contaminated by reviewing all operation records including events and accident reported from the end of facility construction to the present time. Therefore, the process is time-consuming and required manpower to review extensive plant operational data and documents to conduct the HSA. The method is subjective. Therefore, site characteristics should be determined to detail whether the site is contaminated. For solving these problems, the HSA module of COSMOS used the AI model to automatically differentiate sites through the training data. The output from this module is an area classification in the site of nuclear facilities.

After completing HSA, other RSSI surveys should be planned. The

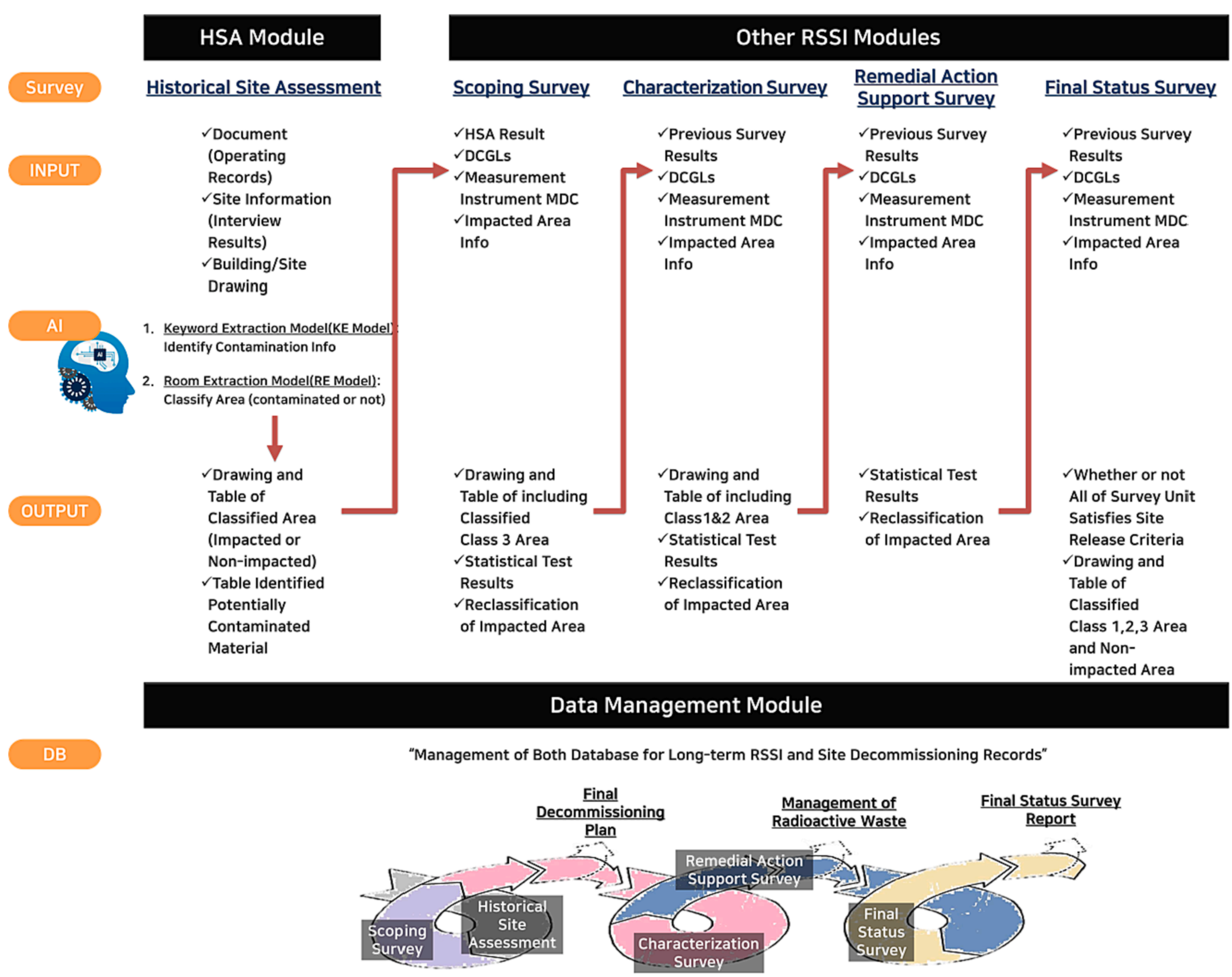


Fig. 1. Software structure of COSMOS.

preliminary survey is conducted to designate the dimension, number, and position of the survey area resulting from the HSA and determine a background area as reference. The results of the preliminary survey would be the input of the management system. Next, survey units are reclassified based on the actual survey data by comparison with DCGLs. The comparison of survey results with DCGLs for the classified survey unit is performed by a statistical method, typically a one-sample statistical test called the sign test or the Wilcoxon Rank Sum (WRS) test reflected in the COSMOS. Other RSSI survey module will perform the statistical test with survey results and provide the results of the reclassification of the survey unit.

Fig. 1 displays the system structure with the input, output, and data management of each module in the COSMOS. The user should input the basic information of the Data Quality Objectives (DQO) process and measured value of radioactive nuclides and should provide survey documents and drawings into the management system. The system processes all user input data to automatically set up the survey unit, calculate the number of data points, and determine the sampling location. The input and output of each survey module is controlled by the data management module during the long-term project. Fig. 2 displays the UI of the COSMOS. In the UI of the COSMOS through data management module, data copying is the role of duplicating files such as user-entered information or documents. It can be shown a blue arrow on display when it is clicked a content to copy and place it in the same number on the next module. And, in order to perform site characteristic assessment, statistical factors such as standard deviation must be calculated by referring to the results of previous surveys. In the COSMOS, we developed the result of one module by connecting it to the next module through the green arrow.

3.2. HSA module

HSA was used to identify potentially contaminated areas by reviewing all available documents. An NPP at the end of OLP should have numerous characteristics documents, including construction records, operational reports, analysis reports, and licensing permission. Therefore, consumer time, and manpower are required to review all data to conduct the HSA. Generally, these data have more than ten thousands of documents and should be reviewed for the identifying area affected by radioactive materials.

Therefore, COSMOS was developed containing two types of the AI model. One type is pre-trained by an unsupervised learning method using document data, and transfer-learned by a supervised learning method using expert evaluation documents. The other type is a framework based on the Deep Neural Network (DNN) to both alleviate the burden of human labor and accelerate the overall execution speed of detection and marking process for drawing data. The first type is the Keyword Extraction (KE) model, and the second type is the Room Extraction (RE) model. The KE model extracts the information of keywords and places for inferring contamination from document data. After objectifying compartments in input drawings, the RE model determines a contaminated place and indicates contamination based on the keyword extraction result of the KE model.

Specifically, in the KE model, we used the Tesseract 5.0 OCR engine based on Long Short-Term Memory (LSTM) models to convert the documents to analytical forms. Text extracted with the Optical Character Recognition (OCR) engine must be preprocessed into word form. Unlike most languages, where words can be parsed according to blank spaces and punctuation marks, certain function words in Korean immediately follow and fuse to nouns and pronouns. Each text component was grammatically annotated by morpheme analysis to detect and distinguish such function words. The analyzed text was parsed into words and placed in a data table according to frequency. (Hokyun Kim et al., 2023). In the RE model, to distinguish each physical compartment, it has detected and objectified each room with the Pixel-wise Object Detection Algorithm (PODA) firstly. And pytesseract detects the number and name for the objectified room and extracts the corresponding characters and numbers that exist in the location detected by OCR. Finally, these extracted results are combined to output objectified room and information about it.

In the KE model, sentence analysis is performed for keyword extraction with document data as inputs. To extract relevant

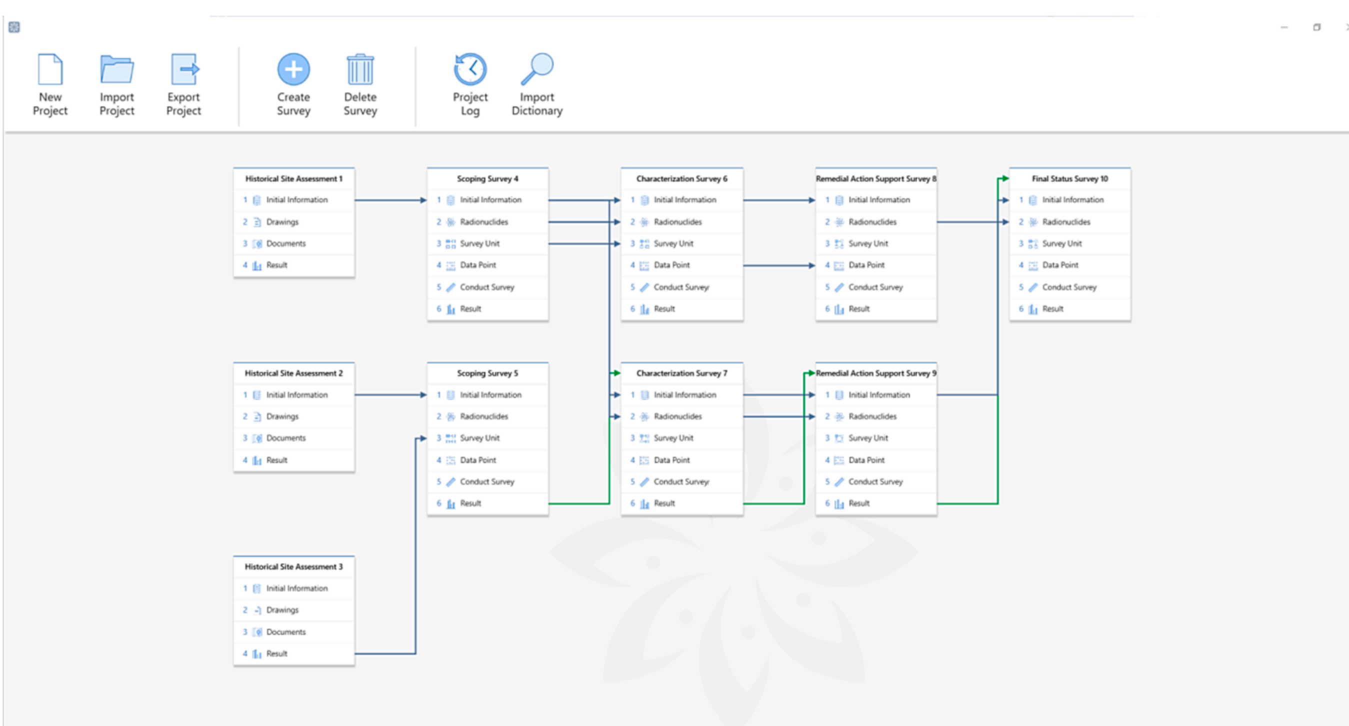


Fig. 2. User interface of COSMOS.

information from queries, in the model, semantic representation are tokenized, and vectorized into numerically processable representations. This process is known as word embedding (S. Lai et al., 2016). The KE model referred to 763 NPPs accident/failure investigation reports provided by the Korea Institute of Nuclear Safety for finding sentences for inferring contaminations and extracting information related to location. Table 2 presents the results of the output from the KE model using reference reports.

The RE model receives input data called room numbering drawing with the name and unique number of each compartment. Inputted drawings require preprocessing before evaluation in the RE model, and COSMOS supports the preprocessing function. The preprocessing of drawing is used for room objectification, which is composed of boundary distinction, name classification, and indication. Fig. 3 displays the preprocessing result on a sample drawing similar to room numbering drawing. In contrast to most DNN-based projects using open-source data, NPP data are prohibited for public use. Therefore, we used a mockup drawing instead of real data of NPPs.

Preprocessed drawings are inputted to the RE model for room objectification. For room objectification, the RE model first determines black dots in drawings. Next, for the center of each black dot, the RE model flows the color in all directions from one dot, and the color does not flows in the same direction on encountering a red line. In this process, a color-filled space with red edge line is separated as a compartment one by one. All pixels in drawings are assigned to the corresponding compartment and the COSMOS prints them. After room separation, the characters covered along the yellow highlight are recognized through around black dot once again, and the RE model extracts the room name and room number from that area. This room objectification process should be performed accurately and sophisticatedly but drawings of NPP are considerably complex. Therefore, COSMOS requires the AI model and preprocessing. The room objectification results of nuclear facility performed by the RE model are correlated with the results of the KE model (i.e., contamination-related keywords) to determine whether the room is contaminated. According to the linkage of each result, this model is visualized, and expressed in the UI of COSMOS.

In addition, the COSMOS can be utilized to perform assessment of contaminated land. Firstly, The KE model allows the extraction of keywords that can infer the area by receiving documents related to contamination of the land, it is the same process as building assessment. In case of the building, room can be classified physically as compartment, but the land does not have a uniform boundary, making it difficult to express the boundary of the contaminated area. To solve this problem, the COSMOS has implemented a function to specify directly contaminated areas on a drawing by user. Then the RE model receives the area of contaminated land indicated by the user, extracts the area along the boundary. Also, the user can enter the name and number of the land directly from result of the RE model. Finally, the contaminated area of the land, which is the result of the RE model, is shown in the drawing through the UI of COSMOS in connection with the results of the KE model.

Table 2
Example of keyword extraction results.*

Report Title	Extracted Sentence	Location	Extracted Keyword (Room Number)
Wolsong Unit 3 reactor coolant leakage by mal-operation of pressurizer drain valve during overhaul	-(18:43:43) according to the first alarm generated in charging pump room (R-403),	Charging pump room	R-403
Reactor manual trip for shielding door maintenance in refueling machine room	Alarm (Completion Check Time Exceeded) has been generated during opening C-side closed door in refueling machine room (R-103) for refueling	Refueling machine room	R-103

* Originally, all data are composed of Korean, but we replicated them as English for understanding.

3.3. Other RSSI modules

The RSSI process is a series of surveys designed to demonstrate compliance with dose or risk-based criteria for sites with residual radioactive (NRC, 2000). The survey design for a site and building is classified by using the survey dataset and potential contamination. The number of samples in each survey unit is calculated by statistical tests using the data collected from the scoping and characterization surveys. The results of the FSS were compared with the release criteria (Hong et al., 2011).

Therefore, COSMOS operates following procedures to prove whether the site is satisfied with the release criteria. First, initial information (e.g., interested nuclides, survey period) for conducting survey, factors for the statistical test and parameters for the calculation of the sampling data point should be set. In COSMOS, 1,278 nuclides are considered to provide options for radioactivity units. A user can input DCGLs for selected interested nuclides and set decision error alpha and beta to evaluate the statistical test. The parameter for the calculation of sampling data point is an area factor that is the magnitude by which the concentration within the small area of elevated activity can exceed DCGLs and maintain compliance with the release criteria. COSMOS can input the area factor of interested nuclides using the import/export function.

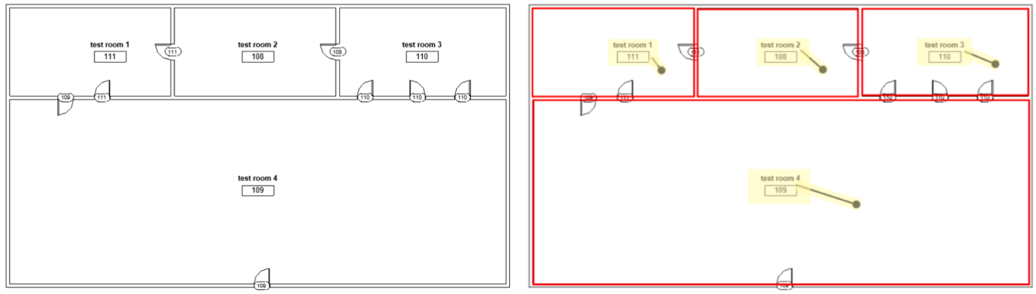
Next, survey units are formed by the class representing the contamination level, referred to the previous survey results. Each survey unit is a physical area consisting of structures or land areas of specified size and shape for which a separate decision is made as to whether or not the area exceeds the release criteria (NRC, 2000). Survey units are formed according to the area standard for each class recommended by the MARSSIM, and the COSMOS has the function of performing survey unit configuration based on the pixel and scale of drawings. Fig. 4 displays an example of a survey unit constructed through the COSMOS. We set arbitrary classes for each room the COSMOS represents colors depending on the contamination level.

Survey units should be evaluated by directly measuring radioactivity or performing sampling analysis. Therefore, sampling data points should be determined with an appropriate approach. The number of data points is calculated by determining whether the background of the survey unit is contaminated. Thus, if contaminants are presented in the background of the survey unit, the background reference area is selected. If a survey unit presents contaminants in the background, then measurements are performed from the reference area, and the survey unit is made using the WRS test. The number of data points is calculated as follows:

$$N = \frac{\left(Z_{1-\alpha} + Z_{1-\beta}\right)^2}{3\left(P_r - 0.5\right)^2} \tag{1}$$

where N is the total number of data points., Z_(1-α)and Z_(1-β) are decision error percentiles and are standard statistical values by the selected decision error levels α and β. P_r is the probability that a random measurement from the survey unit exceeds a random measurement from the background reference area, but less than DCGLs (NRC, 2000).

If the background of survey unit is not contaminated, a background reference area is not necessary. Thus, the survey unit can be evaluated using the sign test that compares directly with the DCGLs. The number of data points is calculated using the following expression for performing the sign test:



(a) Raw data (b) preprocessed drawing

Fig. 3. Result of preprocessed drawing.

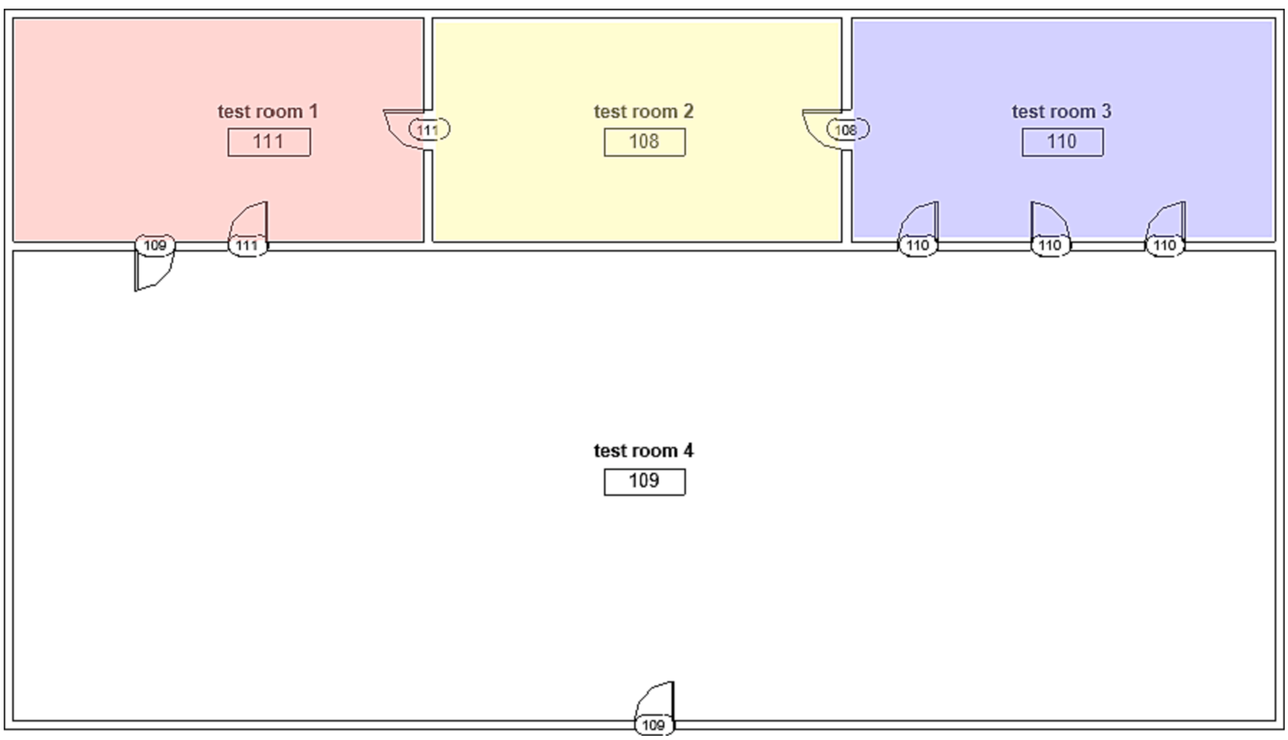


Fig. 4. Example of the survey unit.

$$N = \frac{(Z_{1-\alpha} + Z_{1-\beta})^2}{4(Signp - 0.5)^2} \tag{2}$$

where N, Z_(1-α)and Z_(1-β) are same as Eq. (1). And), Sign p is the estimated probability that a random measurement from the survey unit will be less than the DCGLs. Furthermore, Sign p is used to calculate the minimum number of data points necessary for the survey to meet satisfy the DQOs (NRC, 2000).

After calculating the number of data points through Eq. (1) and Eq. (2), each point can be located either randomly or regularly. For locating and marking them on the, point positions are specified based on the coordinates of the drawing pixel in the COSMOS. For the reference areas, they are randomly positioned through random functions. For other survey units, they are located with a promised direction through the distance between calculated points. Fig. 5. shows displays an marked example of data points in the mockup drawing through the COSMOS.

After marking data points on drawings, users may conduct the characterization survey by direct measurement or sampling analysis. The results of survey are concentrations of each interested nuclides, and the corresponding value is compared with DCGLs to finally perform the statistical test. The WRS test or sign test are conducted by COSMOS, and the conclusion of whether or not the site release criteria are satisfied is detailed.

3.4. Data management module

Notably, MARSSIM advocates for the utilization of nonparametric statistical tests during the creation of FSSs through the DQO process. The objective of this approach is to guarantee that the survey findings exhibit sufficient quality and quantity to effectively determine whether the contamination levels in survey units satisfy DCGLs, particularly for crucial final decisions (NRC, 2000).

As mentioned, site characterization requires overall data management in the evaluation process during the long-term period. The data management module manages the input and output in the HSA and other RSSI modules as the database during the decommissioning project. The module consists of the database and denotes that the result of the specific survey step is automatically and sequentially used as the input of next survey step. This process can minimize human errors, time, and manpower, and ensures the reliability and quality of data management process by preventing risks such as data loss. By collecting and analyzing data on the physical, chemical, and radiological conditions of the

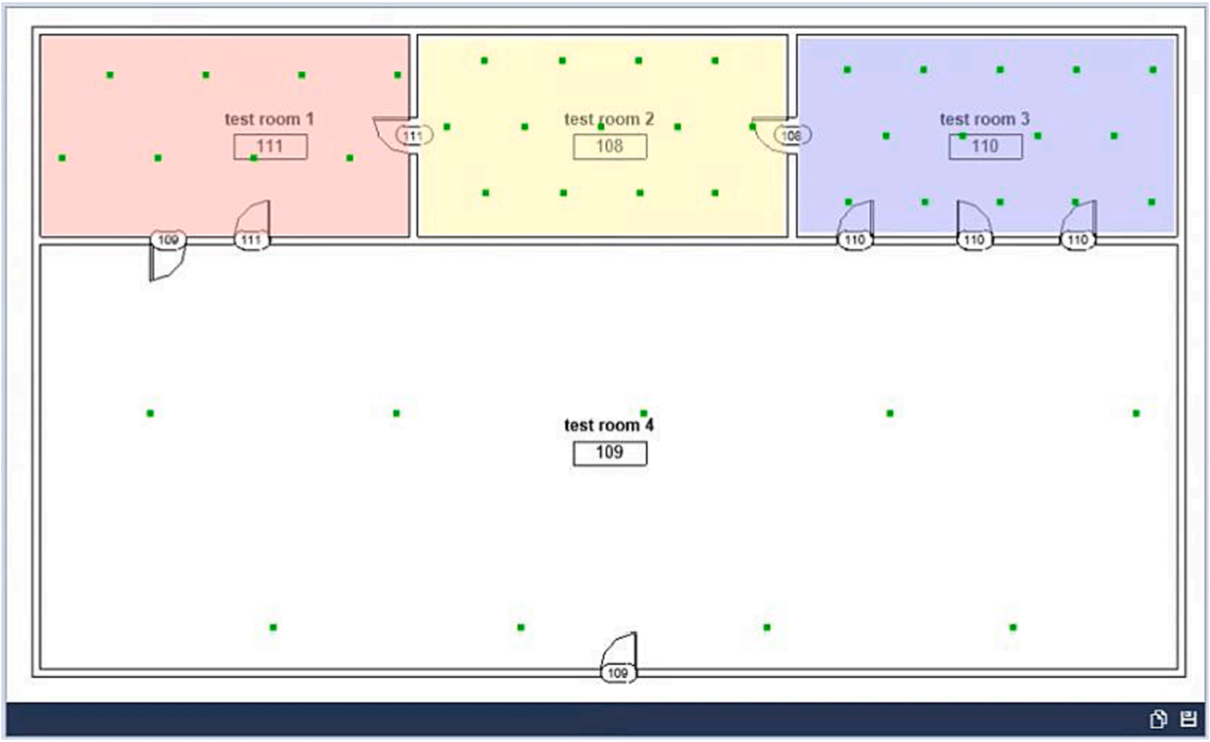


Fig. 5. Example of the data points result.

facilities, site characterization evaluates the extent and type of contamination and also assists in selecting the suitable decommissioning strategy, developing the decommissioning plan, estimating the costs and risks, and handling radioactive, and nonradioactive materials. The data management module can support the organization and coordination of the various tasks, such as sampling, analysis, modeling, reporting and quality assurance, in site characterization, and can enhance the communication and the cooperation among various stakeholders, such as regulators, operators, contractors, and the public. Therefore, Fig. 6 displays COSMOS a database with its data relationships. In entity-relationship diagrams (ERDs), lines and dotted lines are used to represent different types of relationships between entities and their attributes. Solid lines are used to represent strong or definite relationships between entities or entity sets. These relationships indicate that there is a direct and mandatory connection between the entities involved. On the other hand, dotted lines represent weak or optional relationships between entities or entity sets. These relationships indicate that the connection is not mandatory, and one entity may or may not be related to the other.

4. Test data set

As mentioned, the real drawings of NPPs cannot be disclosed because of confidentiality concerns. Therefore, we collected mockup drawings similar to those of NPPs from the web and tested the code. Some room names and numbers were changed intentionally to be consistent with the reports indicating the leakage of radioactive materials. Next, the drawing was preprocessed to be input data for the RE model in the COSMOS. The corresponding drawing has 45 rooms and two rooms are presented to be contaminated as per the reports. Co-60 and Cs-137 were nuclides of interest and DCGLs were assumed to be 1.0 and 0.1 Bq/g, respectively.

For conducting site characterization, any statistical test may be used provided that data are consistent with the assumptions underlying their use. In COSMOS, statistical test models, which can directly compare each survey unit with the applicable release criteria, are used. For contaminants in the background, the WRS test is used. When contaminants are not present in the background, the sign test is used. To determine the number of data points necessary for these tests, acceptable probabilities of making type I decision errors (α) and Type II decision errors (β) were established equally as 0.05.

5. Results and discussion

First, two documents, which are containing sentences expressing that room contamination, were input into the KE model. The KE model read each document and determined to be potentially contaminated rooms. Table 3 presents the results.

A test drawing was preprocessed to input the RE model including boundary distention, name classification, and indication. The RE model extracts all rooms in the drawing and combines them with results of the KE model, which informed a contaminated room. Fig. 7 displays these results to determine the status of rooms contamination.

COSMOS distinguished two contaminated rooms from the total of 45 rooms based on the input documents of KE model in the input drawing. The colored areas indicate contaminated rooms classified as class 1 with red and class 2 with yellow, respectively. Gray areas are uncontaminated rooms. Most rooms in the test case have been excluded from the statistical test because MARSSIM revealed that performing the statistical test on rooms with sufficient evidence can prove that they are not contaminated. However, two rooms were determined to be contaminated, which should be satisfied the requirements of regulatory guidance. A red room was matched with a non-contaminated survey unit to perform the WRS test, and the sign test was used for a yellow room.

Next, the number of data points should be calculated for each survey unit. This phenomenon was derived through the parameter of the statistical test set. Four data points were calculated for the survey unit to perform the WRS test, and five data points were used for the sign test. COSMOS visualizes all calculated data points in the test drawing, and the configuration and distance of data points is according to the setting shape (i.e. triangular or rectangular). Data points are randomly displayed for non-contaminated rooms. Fig. 8 displays the result of test case representing data points as green dots.

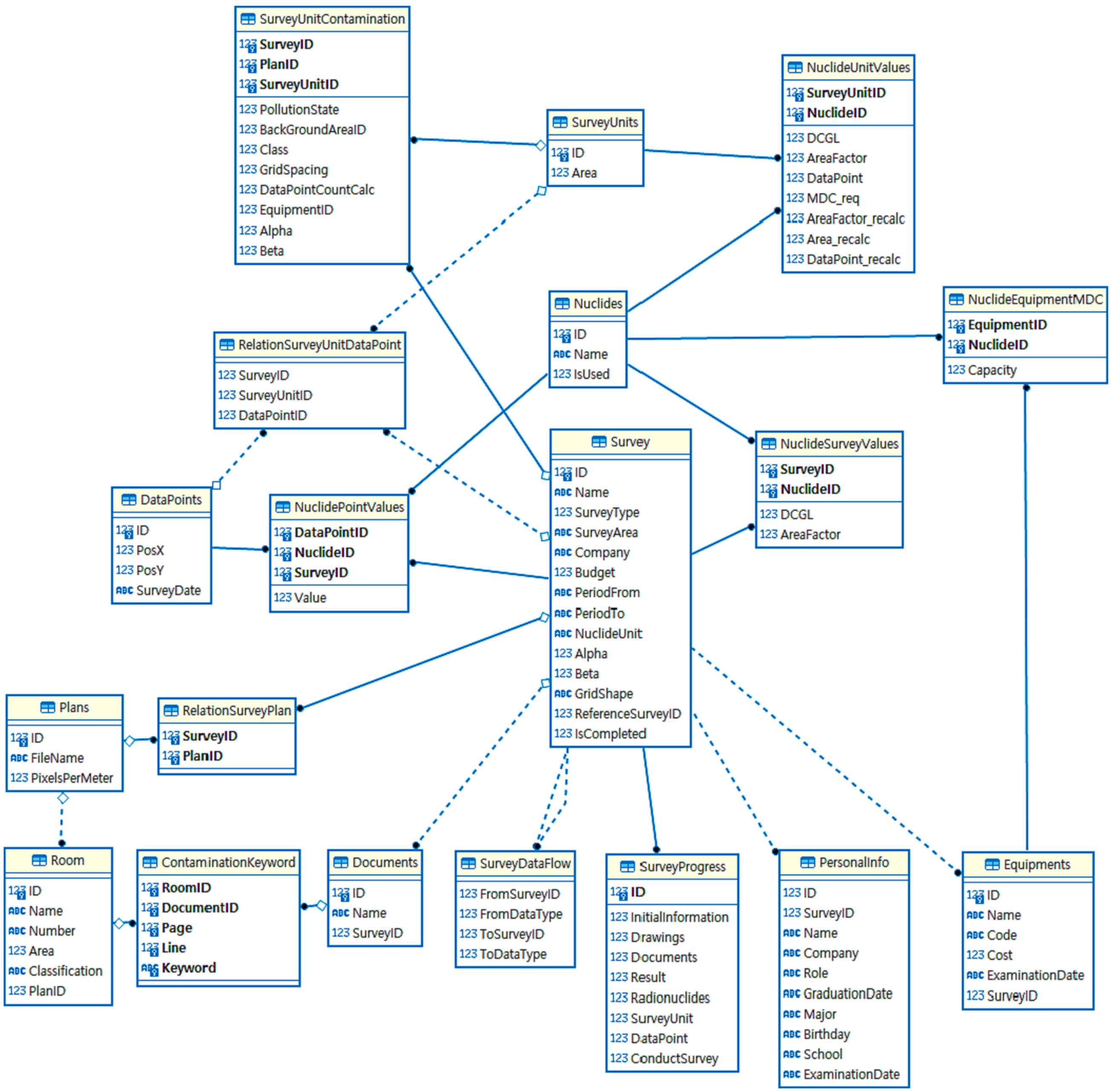


Fig. 6. Entity relationship diagram of database.

Table 3
Keyword extraction result of the test case.*

Report ID	Extracted Sentence	Place Name	Extracted Keyword (Room Number)
2004-07 (040618 W2)	the location of damaged tube is upper area of R-108 the refueling machine room (A-side)	Refueling machine room	R-108
2006-6 (060404 W3)	sharp increase of tritium concentration in the reactor building boiler room (R-501)	Reactor building boiler room	R-501

* Originally, all data are composed of Korean, but we replicated them as English for understanding.

Finally, the test values of the interested nuclides (i.e., Co-60 and Cs-137) for each data point were entered in the COSMOS program, and the statistical test results of each survey unit were derived within seconds after all input. Table 4 presents the result of the statistical test.

The WRS and sign tests were performed to determine whether reject or not to reject the null hypothesis that the survey unit does not satisfy release criteria. The result of the WRS test conducted through the COSMOS reveals that the R-501 room cannot be released because the contamination level is higher than the DCGL. By contrast, the R-108 room reveals that the site release criteria are satisfied through the sign test.

In addition, regarding validation of the COSMOS, it is derived that performance of AI model can replace human's ability rather than comparing it with other existing programs described in section 2 because it is the first software for performing overall site characterization.

Firstly, for comparison with the KE model, one site characterization

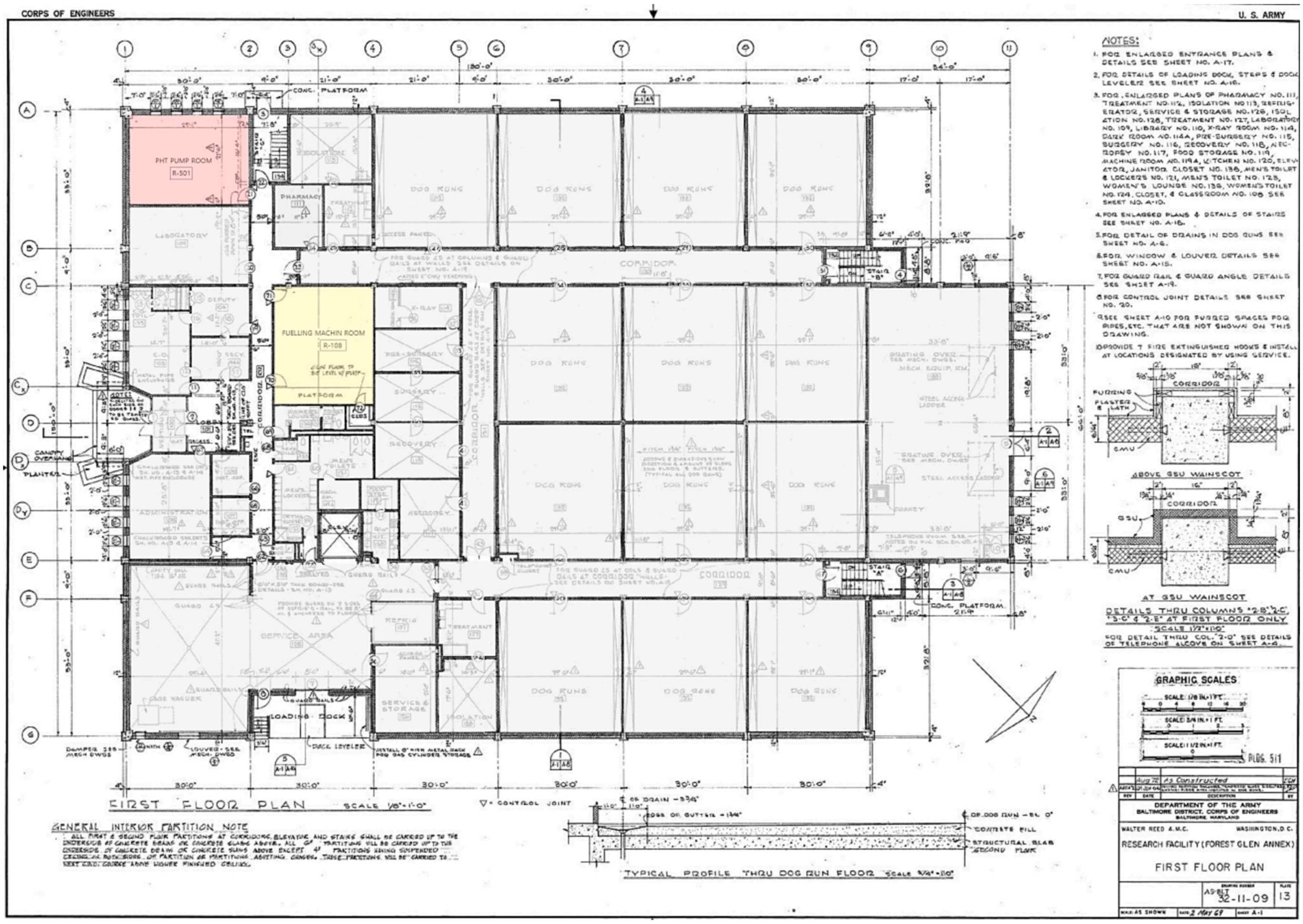


Fig. 7. Room extraction results of test case.

expert and two engineers with no experience of site characterization were selected to extract contamination keywords from 10 documents chosen arbitrarily. These documents are recorded in the event of an accident or failure at NPPs in Korea and contain information about the area if there is a contaminated area. The contaminated area is the area where the accident or failure has clearly caused physical contamination with radioactive nuclides. And process of extracting this contaminated area is the same between KE model and expert, because KE model has developed as a rule-based algorithm which is based on the principle of expert's assessment. As a result, the performance time per document page took 19.2 s for the KE model and 11.1 s for humans on average. However, the accuracy of contamination keyword extraction was 100 % in the KE model and 80 % in humans on average. This can be shown that the KE model was a bit slower than human for small amount of documents, but the actual HSA requires checking thousands of documents at least. Therefore, the KE model capable of 24/7 operation can overcome human limitations to shorten the total performance time and can be extracted more accurately.

The other AI model, the RE model was evaluated with two drawings of each 50 and 4 physically divided compartments. As a result, compartments were detected with 100 % accuracy. However, it was hard to accurately extract the room number and name for low-quality drawing as shown in Fig. 9. The other high-quality drawing was extracted perfectly. In order to solve this problem, we will perform it in future research as it seems to be possible to overcome the image quality improvement of the drawing by using the technique of super resolution (Christian Ledig et al., 2017).

6. Conclusion

We integrated software called COSMOS for site characterization, specifically designed to process the document, and drawing data of NPPs. COSMOS has functions of user input, data processing and assessment, and reporting, and is composed of the HSA, RSSI, and data management modules. The HSA module used the AI model based on unsupervised learning and DNN to identify potentially contaminated areas from documents and drawings. RSSI modules perform the statistical test with measurement data, and the data management module can control the entire process of RSSI evaluation for the long-term project.

We constructed a test case to verify the function of COSMOS for site characterization and described the evaluation results. Two documents were used to extract keywords informing contaminations, and one drawing was input for objectification. The AI model accurately recognized two contaminated rooms among all the rooms classified and subsequently formed survey units to perform the statistical test. COSMOS calculated the number of data points and the distance between points. Data points were marked on the drawing according to the shape to be formed. Finally, the setting value of interested nuclides for each data point was entered into COSMOS and a statistical test comparing with DCGL was performed to show results within seconds.

COSMOS can perform site characterization through statistical tests to assess whether the site satisfies the release criteria. If the result revealed that the release criteria are exceeded or if additional data points are necessary, decision makers would determine appropriate further actions. We anticipated that COSMOS can be used to provide valuable information for the radiation protection in decommissioning projects

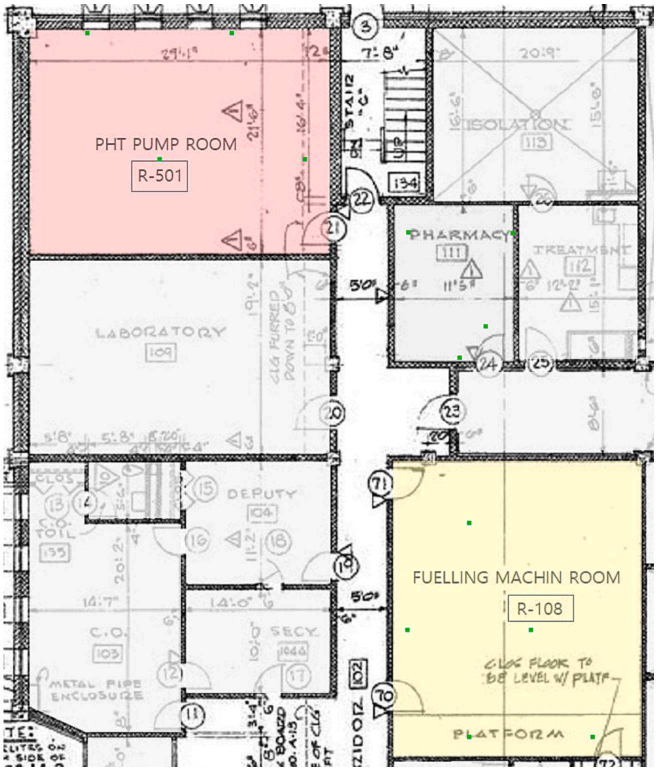


Fig. 8. Data points for each survey unit in the test case.

Table 4
Statistical test results of the test case.

Test Type	Room Number	Room Name	Classification	Result
WRS	R-501	PHT PUMP ROOM	Class 1	Null Hypothesis is accepted.
Sign	R-108	FUELLING MACHINE ROOM	Class 2	Null Hypothesis is rejected.

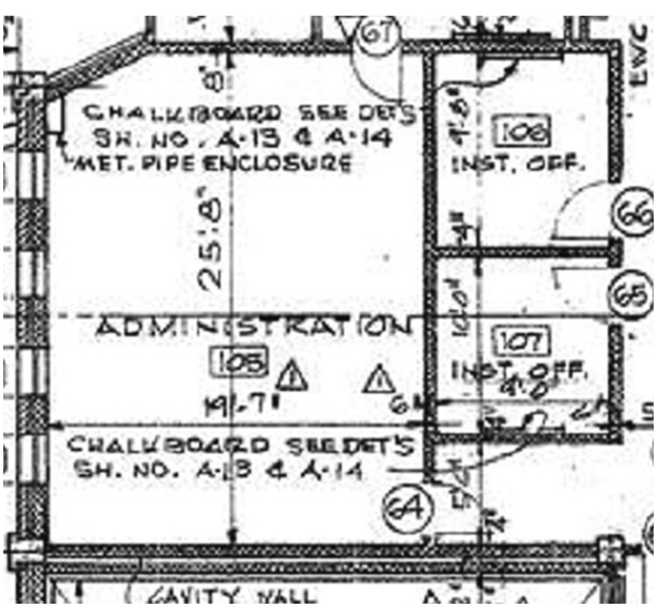


Fig. 9. Example of inaccurately extracted low quality drawing.

and aid in preparing the final decommissioning plan and the FSS report. COSMOS can be used as a supportive tool for site characterization saving resources during the decommissioning project of nuclear facilities.

CRediT authorship contribution statement

Hyungi Byun: Conceptualization, Methodology, Validation, Investigation, Writing – original draft. **Jong Dae Park:** Methodology, Software, Validation, Visualization. **Sihyun An:** Methodology, Software, Validation, Data curation, Writing – original draft. **Jinu Kim:** Methodology, Software, Validation, Data curation, Writing – original draft. **Juyub Kim:** Conceptualization, Methodology, Validation, Writing – review & editing, Supervision. **Doo Yong Lee:** Conceptualization, Writing – review & editing, Project administration, Funding acquisition. **Bongsoo Lee:** Conceptualization, Methodology, Writing – review & editing, Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

This work was supported by the Technology Innovation Program (or Industrial Strategic Technology Development Program - ATC +) (20014125, Development of Intelligent Management Solution for Nuclear Decommissioning Site Characterization) funded By the Ministry of Trade, Industry & Energy (MOTIE, Korea)

References

Radiation Safety Information Computational Center, COMPASS 1.0.0: Computerization of MARSSIM for Planning and Assessing Site Surveys, RSICC CODE PACKAGE PSR-520, 2002.

CSN, Instruction IS-13 on the radiological criteria for the release of nuclear installation sites, Legal dossier III: Technical regulation, 2007.

EPA, Statistical software for environmental applications for data sets with and without nondetect observations, ProUCL Version 5.0.00 Technical Guide, 2013.

IAEA, Selection of decommissioning strategies: Issues and factors, IAEA-TECDOC-1478, 2005.

IAEA, Release of sites from regulatory control on termination of practices, IAEA-WS-G-5.1, 2006.

Kim, H., Park, S., Jeong, M., Byun, H., Kim, J., 2023. Doo Yong Lee, Jooyoung Jeon, Eokin Yi, Kwangwon Ahn, Scailing behavior and text cohesion in Korean texts. PLoS ONE 18 (8), e0290168.

Pacific Northwest National Laboratory, Visual sample plan version 1.0 user's guide, Interim Report, 2001.

Lai, S., Liu, K., He, S., Zhao, J., 2016. How to generate a good word embedding. IEEE Intell. Syst. 31 (6), 5–14.

Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, Wenzhe Shi, Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network, *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.

NRC, Multi-agency radiation survey and site investigation manual, NUREG-1575, 2000.

NRC, User's manual for RESRAD-OFFSITE code version 4, NUREG/CR-7268, 2020.

NRC, Radiological criteria for unrestricted use, 10 CFR 20 Subpart E 20.1402, 2021.

OECD/NEA, Nuclear site remediation and restoration during decommissioning of nuclear installations, NEA No. 7192, 2014.

Nuclear Safety and Security Commission, Criteria for reuse of site and buildings after completion of decommissioning of nuclear facilities, (Nuclear Safety and Security Commission No. 2021-11), 2021.

Sang Beom Hong, Kiwon Lee, Jinho Park, Unsoo Chung, Application of MARSSIM for Final Status Survey of the Decommissioning Project, *J. Nucl. Fuel Cycle Waste Technol.*, vol. 9, pp. 107-111, 2011.

Robert Stewart, Chris Welsh, Tom Purucker, An introduction to spatial analysis and decision assistance (SADA), Environment Application for Version 5 User Guide, 2009.