EuroSDR

European Spatial Data Research

February 2023

Survey on Machine Learning / Deep Learning Projects

Results 2021

Tobias Kellenberger, André Streilein, Joep Crompvoets, Fabio Remondino

Official Survey Report

Published by EuroSDR

EUROPEAN SPATIAL DATA RESEARCH

PRESIDENT 2020 – 2023:
Michael Hovenbitzer, Germany

VICE-PRESIDENT 2021 – 2023:
Fabio Remondino, Italy

SECRETARY – GENERAL 2019 – 2023:
Joep Crompvoets, Belgium

DELEGATES BY MEMBER COUNTRY:
Austria: Wolfgang Gold, Gottfried Mandlburger
Belgium: Eric Bayers
Croatia: Željko Bačič
Cyprus: Elikkos Elia, Georgia Papathoma, Andreas Hadjiraftis, Dimitrios Skarlatos
Denmark: Jesper Weng Haar, Tessa Anderson
Estonia: Tambet Tiits, Artu Ellmann, Evelyn Uuemaa
Finland: Juha Hyyppä, Juha Kareinen
France: Bénédicte Bucher, Aurélien Plyer
Germany: Michael Hovenbitzer, Norbert Haala
Ireland: Paul Kane, Audrey Martin
Norway: Jon Arne Trollvik, Ivar Maalen-Johansen
Poland: Anna Bober, Krzysztof Bakuła
Portugal: Mário Caetano, Rui Reis, Marisa Silva
Slovenia: Dalibor Radovan, Peter Prešeren, Marjan Čeh
Spain: Julián Delgado Hernández
Sweden: Tobias Lindholm, Anders Rydén, Heather Reese
Switzerland: André Streilein
The Netherlands: Martijn Rijsdijk
United Kingdom: Claire Ellul

ASSOCIATE MEMBERS AND THEIR REPRESENTATIVES:
Digitaal Vlaanderen: Jo Van Valckenborgh
Esri: Nick Land, Konrad Wenzel (nFrames)
Field Group: Leif Erik Blankenberg
Hexagon: Simon Musaeus
ICGC: Julià Talaya
IGI: Philipp Grimm
RIEGL: Peter Rieger
Vexcel: Michael Gruber
1Spatial: Dan Warner

COMMISSION CHAIRPERSONS:

Data Acquisition: Jon Mills, United Kingdom
Modelling, Integration and Processing: Norbert Haala, Germany
Information Usage and Visualization: Martijn Rijsdijk, The Netherlands
Business Models and Operations: Frédéric Cantat, France
Knowledge Transfer: Anka Lisec, Slovenia

OFFICE OF PUBLICATIONS:

Bundesamt für Eich- und Vermessungswesen
Publications Officer: Christine Ressl
Austria
Contact via EuroSDR secretariat

CONTACT DETAILS:

EuroSDR Secretariat
KU Leuven Public Governance Institute
Faculty of Social Sciences
Parkstraat 45 bus 3609
3000 Leuven
Belgium

Tel.: +32 16 37 98 10
Email: eurosdr@kuleuven.be
Web: www.eurosdr.net

The official publications of EuroSDR are peer-reviewed.

Tobias Kellenberger, André Streilein, Joep Crompvoets, Fabio Remondino

"Survey on Machine Learning / Deep Learning Projects"
Results 2021

Table of contents

Index of Figures 6
1 Introduction 7
2 Methodology 7
2.1 Machine Learning (ML) 7
2.2 Deep Learning (DL) 8
3 Results 9

SURVEY ON MACHINE LEARNING / DEEP LEARNING PROJECTS

Results 2021

With 12 figures

Tobias Kellenberger [a], André Streilein [a],
Joep Crompvoets [b], Fabio Remondino [c]

[a] Federal Office of Topography swisstopo
Seftigenstrasse 264, 3084 Wabern, Switzerland

[b] KU Leuven Public Governance Institute
Parkstraat 45, 3000 Leuven, Belgium

[c] Bruno Kessler Foundation
via Sommarive 18, 38123 Povo-Trento, Italy

Index of Figures

Figure 1: Questionnaire form. 8

Figure 2: Respondents' countries. 9

Figure 3: Countries with reported projects. 9

Figure 4: Single-word cloud highlighting the frequency and relevance of ML/DL-project objectives. 10

Figure 5: Collocating words cloud highlighting the frequency and relevance of ML/DL-project objectives. 10

Figure 6: Purposes of Interest in using ML/DL. 11

Figure 7: Project status and comparison with 2018 results. 11

Figure 8: Project contractor/research entity and comparison with 2018 results. 12

Figure 9: Type of algorithm used in the project. 12

Figure 10: Type of Machine Learning algorithm used in the project. 13

Figure 11: Collocating-words clouds highlighting the frequency and relevance of input geodata and auxiliary data. 13

Figure 12: Implemented in production-line and comparison with 2018. 14

1 Introduction

The adoption of Artificial Intelligence (AI) at National Mapping and Cadastral Agencies (NMCAs) in Europe could bring benefits such as personalised services for citizens, fewer repetitive tasks for staff members (liberating time for more tasks that have more value as listening, creating, improving services, …), detect harmful content, support decision-making processes where a high number of data and variables are involved, which is more and more frequent when societal challenges are involved.

However there has been very limited initiatives aiming at measuring and understanding the level of AI adoption by NMCAs in Europe. As a follow up action of the successful Joint Virtual Workshop organized by EuroSDR in collaboration with EuroGeographics on "Artificial Intelligence for NMCAs" held on 03-04.02.2021, the Executive Management Team of EuroSDR decided to repeat the EuroSDR survey on Machine Learning / Deep Learning conducted in 2018. EuroSDR received valuable support of swisstopo to successfully launch this survey.

Although this is one of the first attempts to draw the landscape of the use of AI at NMCAs in Europe, the survey provides a unique overview of the current use of AI at NMCAs in Europe – insights which are extremely lacking at this stage and some recommendations to support AI adoption in order to improve their services.

2 Methodology

The survey was open for all NMCAs across Europe interested in the topic of Artificial Intelligence. The survey consisted of 16 questions (closed and open). It was open from 15 March 2021 to 15 June 2021. The questions were related to possible projects at NMCAs in the technical field of Machine Learning / Deep Learning. The project may be completed, ongoing or even planned. In case of having more than one project to report, NMCAs were asked to use the survey form multiple times.

The following definitions (from Wikipedia) were given for Machine Learning and Deep Learning.

2.1 Machine Learning (ML)

Machine Learning (ML): The field of AI that uses statistical techniques to give computer systems the ability to give computer systems the ability to "learn" (e.g., progressively improve performance of a specific task) from data, without being explicitly programmed. Machine Learning tasks are typically classified into several broad categories:

- Supervised learning: The computer is presented with example inputs and their desired outputs, given by a "teacher", and the goal is to learn a general rule that maps inputs to outputs. As special cases, the input signal can be only partial available, or restricted to special feedback.
- Semi-supervised learning: The computer is given only an incomplete training signal: a training set with some (often many) of the target outputs missing.
- Active learning: The computer can only obtain training labels for a limited set of instances (based on a budget), and also has to optimize its choice of objects to acquire labels for. When used interactively, these can be presented to the user for labelling.
- Unsupervised learning: No labels are given to the learning algorithm, leaving it on its own to find structure in it input. Unsupervised learning can be a goal in itself (discovering hidden patterns in data) or a means towards an end (feature learning).
- Reinforcement learning: Data (in form of rewards and punishments) are given only as feedback to the program's actions in a dynamic environment, such a driving a vehicle or playing a game against an opponent.

2.2 *Deep Learning (DL)*

Deep Learning (DL): is a class of Machine Learning algorithms that use a cascade of multiple layers of nonlinear processing units for feature extraction and transformation. Each successive layer uses the output from the previous layer as input. Deep Learning could learn in supervised (e.g., classification) and/or unsupervised (e.g., pattern analysis) manners. It could learn from multiple levels of representations that correspond to different levels of abstraction: the levels form a hierarchy of concepts.

The questionnaire had the following form to report one project. If more than one project could be reported, then use the form multiple times and send back one completed form for each project.

The form consisted of two main parts: 1) Organization and 2) Project.

Organisation

Organisation	Click or tap here to enter text.
Country	Click or tap here to enter text.
Responsible Person(s)	Click or tap here to enter text.
Postal Address	Click or tap here to enter text.
Email Address	Click or tap here to enter text.
Contact Phone	Click or tap here to enter text.

Project

Name of the Project	Click or tap here to enter text.
Short description	Click or tap here to enter text.
Three main questions to be answered in the project	1. Click or tap here to enter text. 2. Click or tap here to enter text. 3. Click or tap here to enter text.
Project time frame	☐planned ☐running ☐completed Start date: Click or tap to enter a date. End date: Click or tap to enter a date.
Project contractor / research entity	☐internal ☐external ☐both
Type of Machine Learning	☐Supervised learning ☐Semi-supervised learning. ☐Active learning ☐Unsupervised learning ☐Reinforcement learning
Input data type	Geodata: Click or tap here to enter text. Auxiliary data: Click or tap here to enter text. Others: Click or tap here to enter text.
Pre-requisites	Click or tap here to enter text.
Main findings	Click or tap here to enter text.
Implemented in production-line	☐ yes ☐ no

Figure 1: Questionnaire form.

3 Results

38 responses from mapping agencies in 20 countries (Austria, Belgium, Denmark, Estonia, Finland, France, Greece, Ireland, Italy, Latvia, Netherlands, Norway, Portugal, Romania, Slovenia, Spain, Sweden, Switzerland, Turkey, United Kingdom) were received. Figure 2 presents a map from where the respondents came from. The respondents came mainly from Western and Northern Europe.

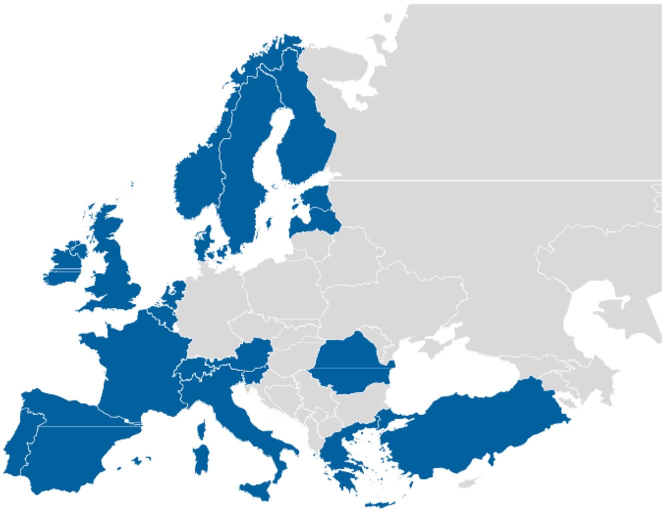


Figure 2: Respondents' countries.

Twenty respondents from 16 organizations in 14 countries (Belgium, Denmark, Finland, France, Greece, Ireland, Italy, Netherlands, Norway, Portugal, Slovenia, Spain, Sweden, Switzerland) reported 32 ML/DL projects. Figure 3 presents the countries in which ML/DL-projects were reported. Most reported ML/DL-projects are from Western and Northern Europe.

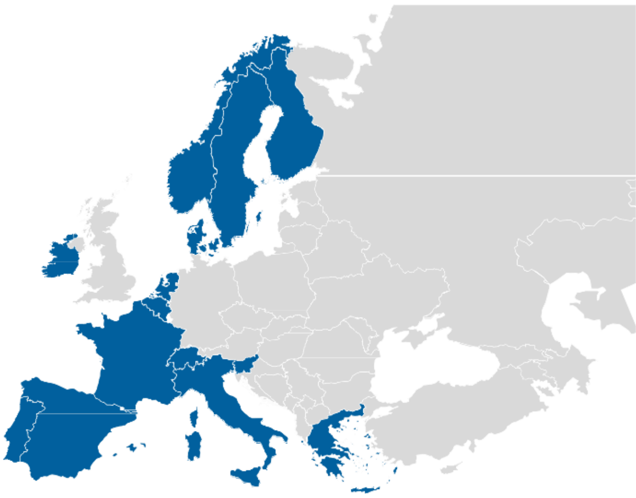


Figure 3: Countries with reported projects.

The main objectives that the reported ML/DL-project aims to achieve are presented in a single word cloud (Figure 4). The figures highlight the frequency and relevance of the objectives. Single words that stand out are *detection, land, map, data, automatically, change, buildings*, *information* and *learning*.

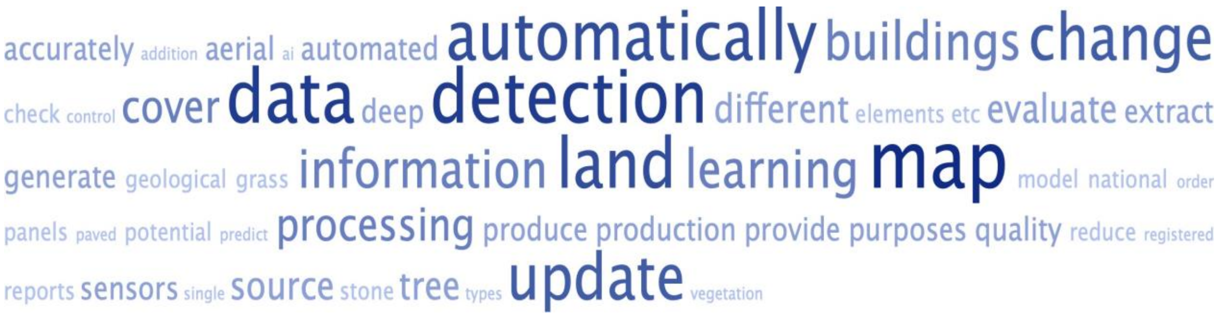


Figure 4: Single-word cloud highlighting the frequency and relevance of ML/DL-project objectives.

Figure 5 presents the ML/DL-projects objectives as collocating words cloud. The objectives that stand out refer to: *Deep Learning, different age, geological report, map update, potential of AI, source data*, and *aerial sensor*.

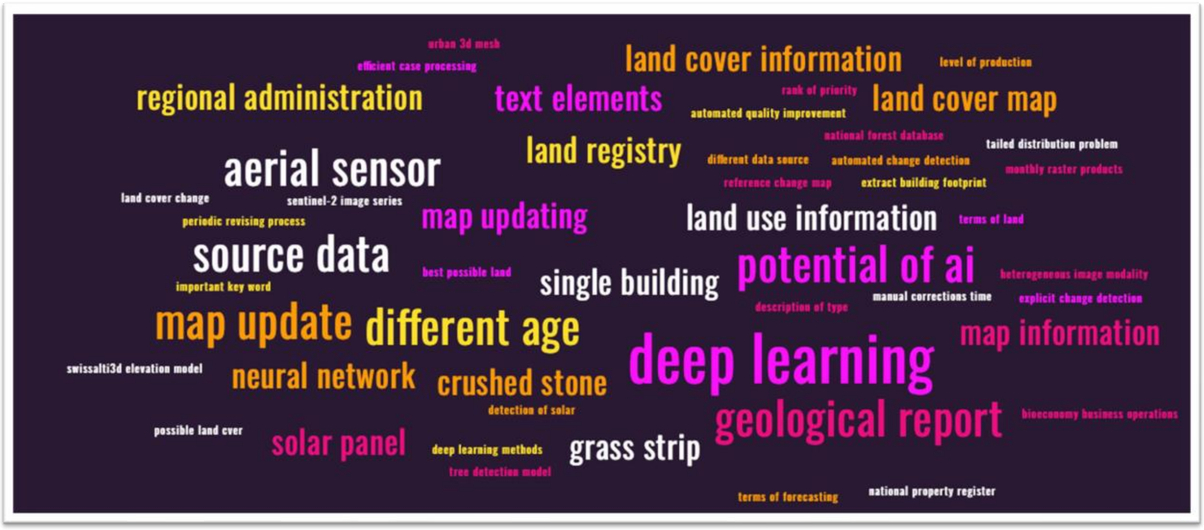


Figure 5: Collocating words cloud highlighting the frequency and relevance of ML/DL-project objectives.

The next figure 6 illustrates for what purposes there was interest in using ML/DL in the reported project. The main intentional purpose of using ML/DL in the project referred to *cost/time savings, more efficient business processes, improved geospatial information quality*, and *development of innovative products/services*. In addition, *expanded use of NMCA data* and *improved delivery of tailored services* were also frequently mentioned as a purpose for using ML/DL.

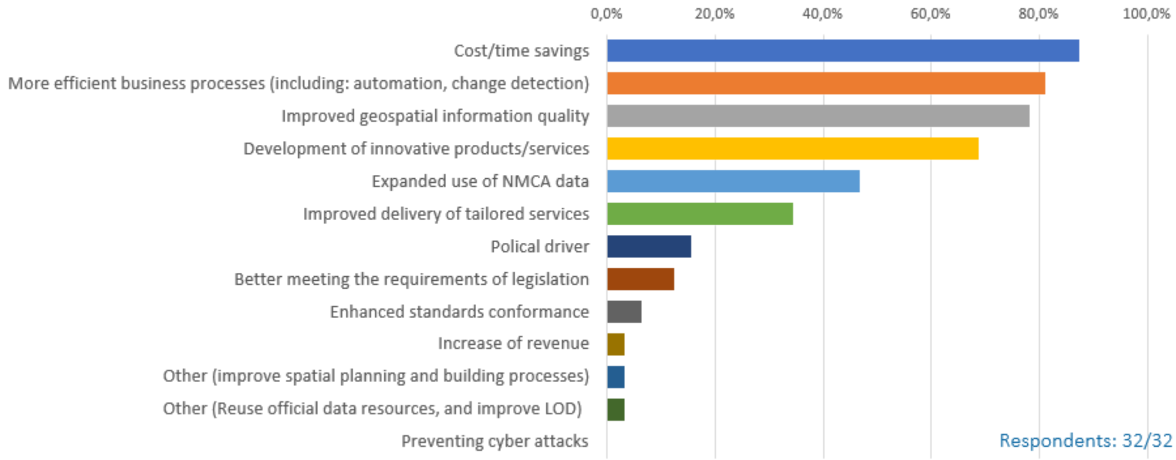
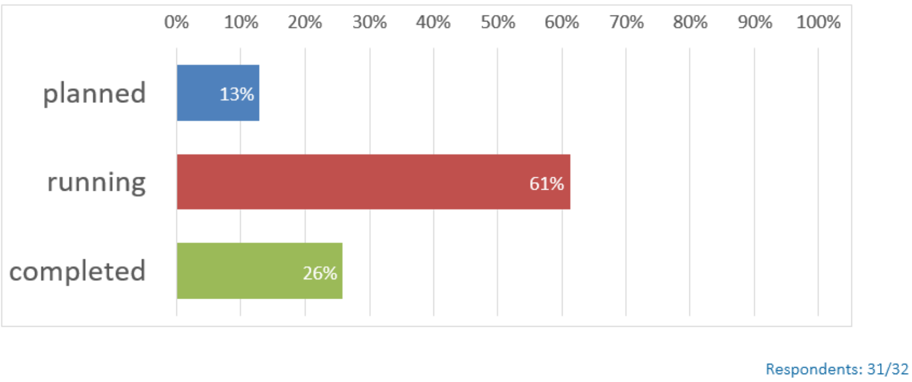


Figure 6: Purposes of Interest in using ML/DL.

When looking to the status of the ML/DL-projects, it appears that most projects were *running* and several projects were *completed*. In comparison with the 2018 survey results, it appears that more projects were completed in 2021 (see Figure 7).



Survey 2021 | **Survey 2018**

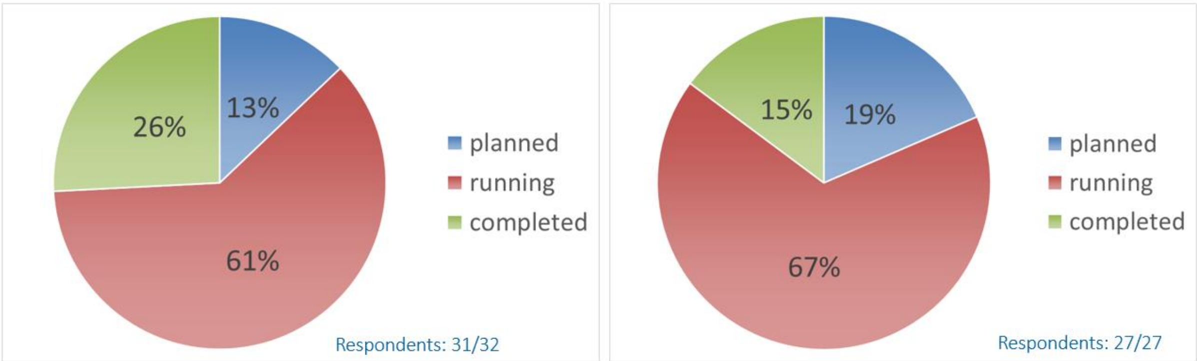
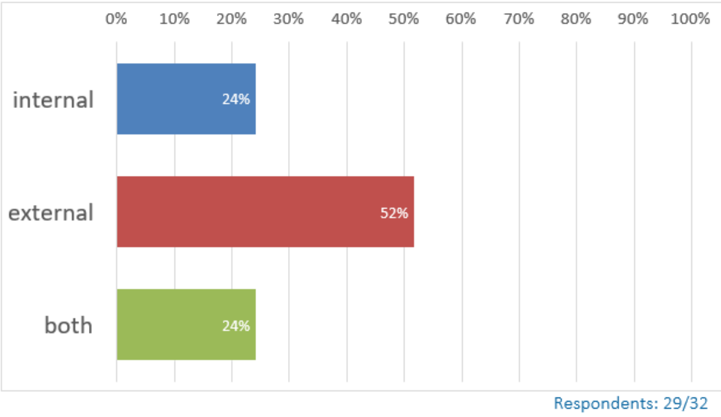


Figure 7: Project status and comparison with 2018 results.

The respondents were also asked about the project contractor/research entity. It appears that most project have an *external* project contractor/research entity (Figure 8). In comparison with 2018 results, a decrease in *internal* and an increase in *both* are noticeable.



Survey 2021 **Survey 2018**

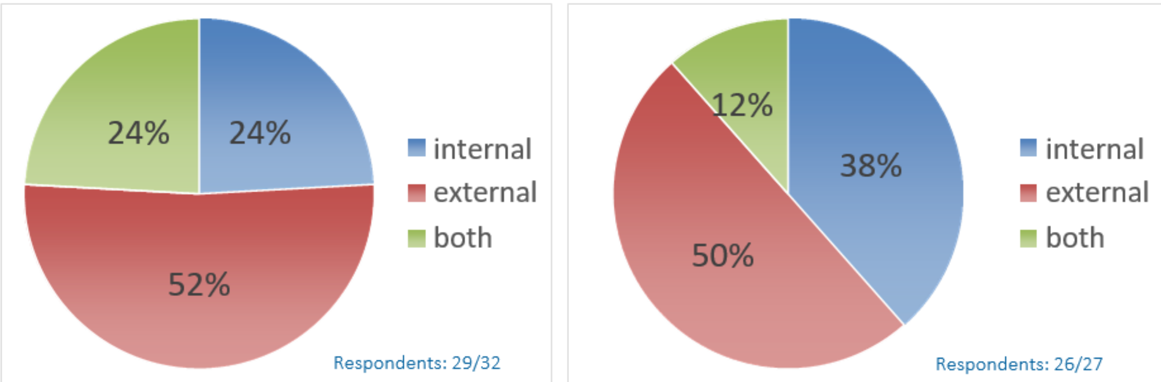


Figure 8: Project contractor/research entity and comparison with 2018 results.

The next question referred to type of algorithm used in the project (Figure 9). It appears that most projects used *a specific Deep Learning algorithm* in the projects. Only a few projects used a *general Machine Learning algorithm*. Other types were occasionally mentioned.

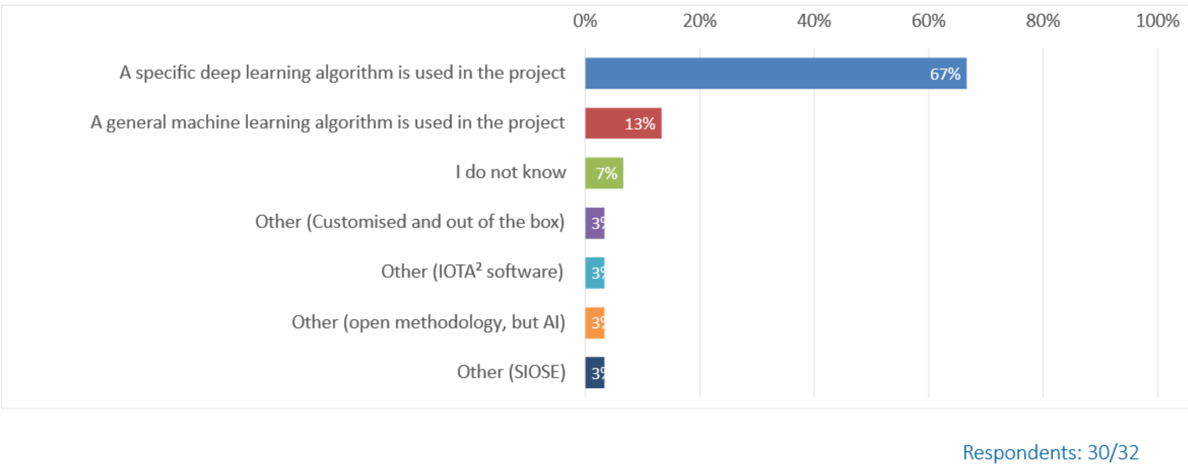


Figure 9: Type of algorithm used in the project.

Another question referred to the type of Machine Learning algorithm used in the project. Figure 10 clearly illustrate that *supervised learning* is the most used type of ML algorithm. *Reinforcement learning, semi-supervised learning* and *unsupervised learning* are just occasionally used.

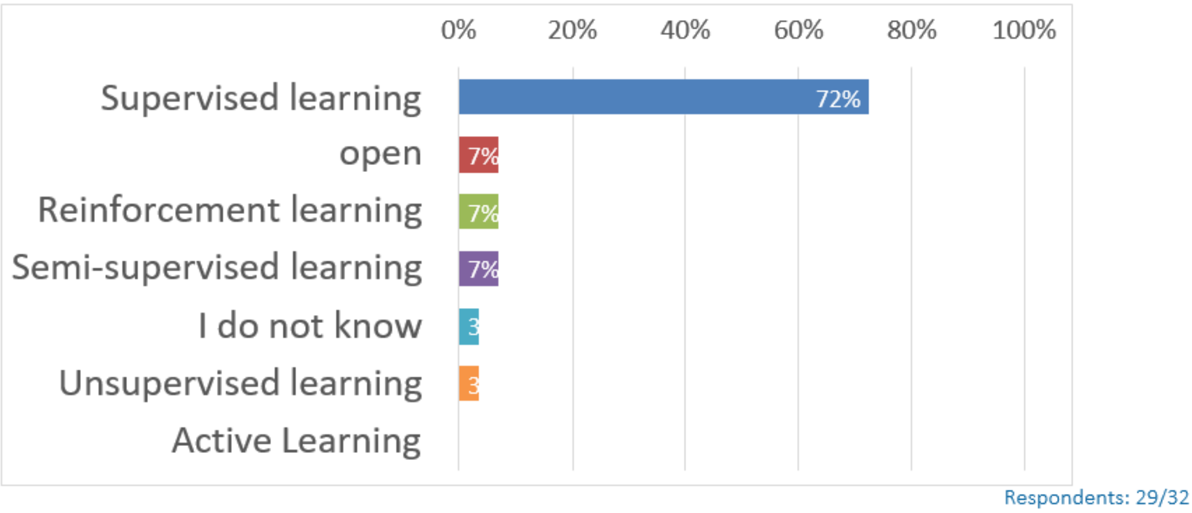
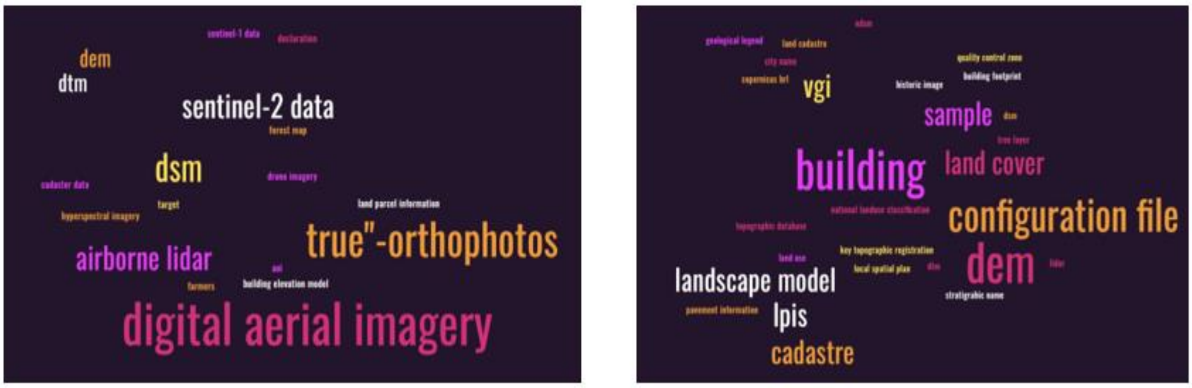


Figure 10: Type of Machine Learning algorithm used in the project.

The next question dealt with input data set(s) for the project. A distinction was made between geodata and auxiliary data. Figure 11 present collocating-words clouds for geodata and auxiliary input data. The main input geodata used are *digital aerial imagery* and *"true"-orthophotos*. Other geodata mentioned were *Sentinel-2 data, DSM*, and *airborne LIDAR*. The main auxiliary data used are *building, configuration file* and *DEM*. Other auxiliary data mentioned were *landscape model, IPIS, sample, land cover* and *cadastre*.

Geodata **Auxiliary data**



Respondents: 29/32

Figure 11: Collocating-words clouds highlighting the frequency and relevance of input geodata and auxiliary data.

The respondents were also asked about the AI-implementation in production-line. It appears that most AI-implementations in production-line are *planned* and a significant number of NMCAs have implemented AI in their production-lines (Figure 12). In comparison with 2018, the number of AI-implementations have been increased and the number of *no* implementations at NMCAs have been dramatically shrunken. This indicates that AI-implementations are becoming more accepted in production-lines of NMCAs.

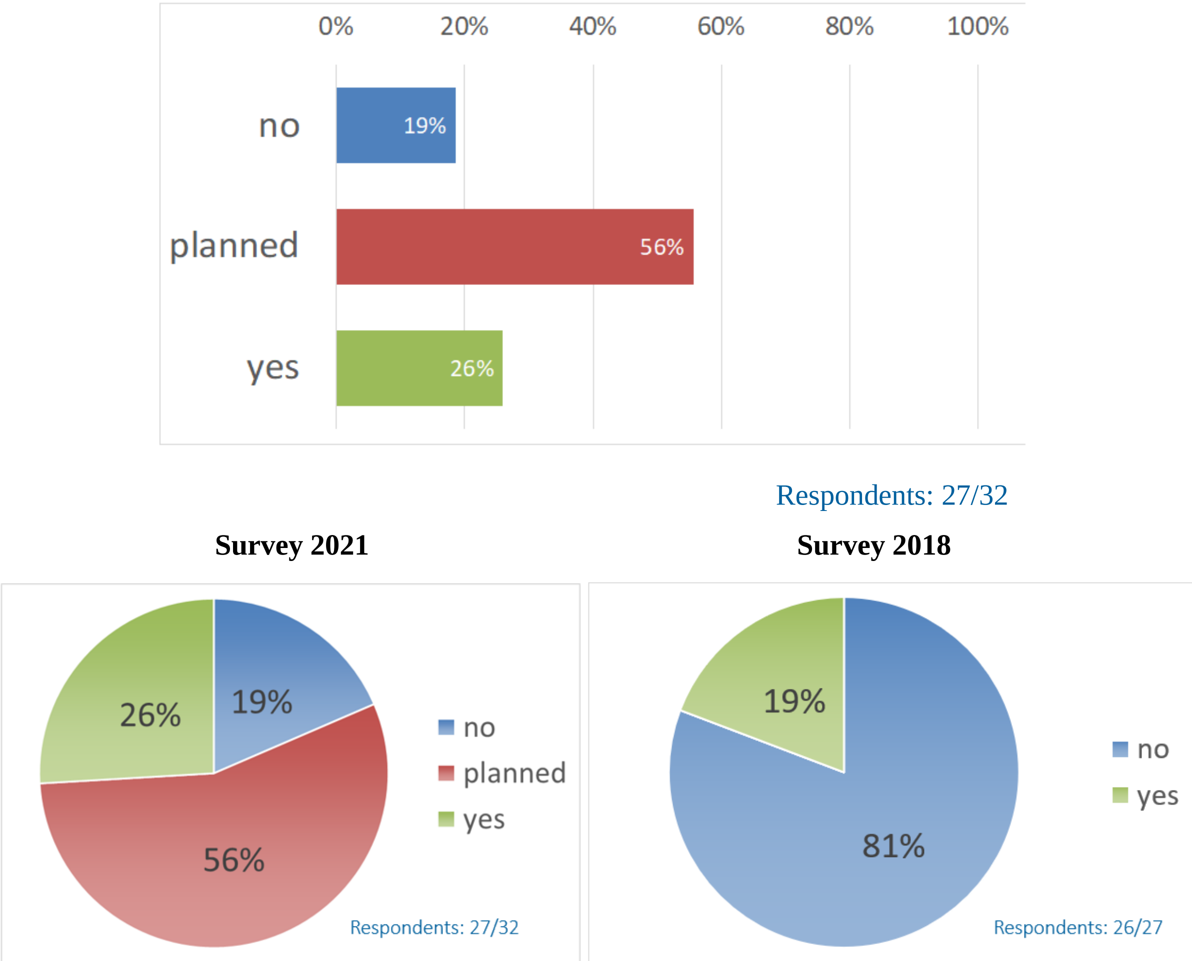


Figure 12: Implemented in production-line and comparison with 2018.

Finally, the main findings from the projects were asked. The list below presents an overview of all the responses given by 29 respondents. The ones in bold were mentioned more than once. The long list and the diversity of findings are remarkable.

List of main findings from the reported projects:

- *Annual land cover map provides insufficient thematic accuracy*
- ***Post-classification analysis based on expert-knowledge is needed***
- *Some problems with smaller buildings*
- *Optimization of resources in time and responsiveness to end users*

- ***Post control quality with semi-manual process is required***
- *Huge effort on labelling*
- *Training data needs to be improved*
- *One model is not sufficient to capture all features*
- *With limited training good accuracies have been achieved*
- ***Customized post-processing needs to be incorporated***
- *It is still hard to figure what the model exactly learns in terms of weights and biases associated with the layers*
- *Discrepancy between historical databases and relevant training data*
- *Need to integrate in larger pipelines*
- *Importance of frequent feedback from thematic experts*
- *A good use of public investments to improve citizen's health*
- ***As expected, we cannot and should not replace human annotators***
- *Already reached 88% accuracy without much hyper parameter tuning*
- *Hard to exclude false positives*
- *Cost saving in the creation of detailed database, in comparison with traditional approaches of photointerpretation*
- ***automatically generated results always need to be checked for plausibility by a human being***