

INSPIRE standards as framework for artificial intelligence applications: a ~~landslides~~ landslide example

Gioachino Roberti¹, Jacob McGregor¹, Sharon Lam¹, David Bigelow¹, Blake Boyko¹, Chris Ahern¹, Victoria Wang¹, Bryan Barnhart¹, Clinton Smyth¹, David Poole^{1,2}, and Stephen Richard¹

¹Minerva Intelligence Inc., 301 – 850 West Hastings Street, Vancouver, British Columbia, Canada V6C 1E1

²Computer Science department, University of British Columbia, Vancouver, V6T 1Z4, Canada

Correspondence: Gioachino Roberti (groberti@minervaintelligence.com)

Abstract. This study presents a landslide susceptibility map using an artificial intelligence (AI) approach ~~that is~~ based on standards set by the INSPIRE framework. ~~We show how INSPIRE standards enhance the interoperability of geospatial data, and enable deeper knowledge development for their interpretation and explainability in AI applications.~~ INSPIRE INSPIRE is a European Union Spatial Data Infrastructure (SDI) initiative to standardize spatial data across borders to ensure interoperability for management of cross-border infrastructure and environmental issues. ~~Despite~~ However, despite the theoretical effectiveness of the SDI, ~~very~~ few real-world applications make use of INSPIRE standards. In this study, we show how INSPIRE standards enhance the interoperability of geospatial data, and enable deeper knowledge development for their interpretation and explainability in AI applications. We designed an ontology of landslides, embedded with INSPIRE vocabularies and then aligned geology, stream network and land cover data sets covering the Veneto region of Italy to the standards. INSPIRE was formally extended to include an extensive landslide type code list, a landslide size code list and the concept of landslide susceptibility to describe map application inputs and outputs. Using the terms in the ontology, we defined conceptual scientific models of ~~slopes~~ areas likely to generate different type of landslides as well as map polygons representing ~~real slopes~~ the land surface. Both landslide models and map polygons were encoded as semantic networks and, by qualitative probabilistic comparison between the two, a similarity score was assigned. The score was then used as a proxy for landslide susceptibility and displayed in web map application. The use of INSPIRE-standardized vocabularies in ontologies that express scientific models promotes the adoption of the standards across the European Union and ~~beyond~~ globally. Further, this application facilitates ~~the explainability of the~~ explaining the generated results. We conclude that public and private organisations, within and outside the European Union, can enhance the value of their data by bringing them into INSPIRE-compliance for use in AI applications.

Copyright statement. This work is distributed under the Creative Commons Attribution NonCommercial-ShareAlike 4.0 International

20 <https://creativecommons.org/licenses/by-nc-sa/4.0/>

1 Introduction

1.1 INSPIRE

Data accessibility and interoperability is key for multinational cross-border applications and fundamental for economic development (European Parliament and the Council, 2007). Different countries have different languages and data standards, hindering infrastructure planning, disaster risk reduction initiatives, and effective legislative implementation. To overcome these challenges, the European Union initiated INSPIRE (Infrastructure for Spatial Information in the European Community - Directive 2007/2/EC ~~-(European Parliament and the Council, 2007)-~~) (European Parliament and the Council, 2007). INSPIRE is structured in 34 spatial data themes organized in three annexes. The themes span ~~social~~-administrative (e. g. street addresses) and environmental domains (e. g. geology), and all EU countries are mandated by law to have implemented the data framework by 2021 (European Parliament and the Council, 2014). Each theme defines a data model and has adopted a set of vocabularies to populate interoperable datasets based on that data model. EU countries are aligning and serving INSPIRE data at a slow pace, and currently relatively few INSPIRE-compliant data sets are available across Europe (Cho and Cromptoets, 2019). Conferences and competitions are currently being organized to promote its implementation and to show the potential impact of real-world applications built on INSPIRE data sets (European Commission, 2019). This project was first presented at one of these conferences, the Helsinki 2019 INSPIRE data challenge under the “Let’s make the most out of INSPIRE!” topic, where the project won first prize.

1.2 Artificial intelligence

Artificial Intelligence (AI) studies "the synthesis and analysis of computational agents that act intelligently" (Poole and Mackworth, 2017). Part of acting intelligently is building models of the world that make predictions. Probabilistic predictions are the most useful ones for subsequent decision making, and can be learned from data (Pearl, 1988). All models are based on human knowledge and data (observations of the world). For some problem domains, society has collected an overwhelming amount of data and still, useful human knowledge of the domain can be very vague. Machine learning has made great progress recently for such cases, particularly with deep learning (Goodfellow et al., 2016). However, for domains with relatively limited, but still very large in volume, data, human knowledge (which may be represented in computer through the use of ontologies) can complement the data to make useful predictions (Pearl, 1988). Many environmental problems do not have enough data (e.g. lack of extensive landslide databases) to be solved by deep learning, but do have enough data to generate useful products when combined with human expertise ~~-(expressed in ontologies)~~ (Poole and Mackworth, 2017). The term Artificial Intelligence is commonly used to indicate only the machine learning part of the field, especially in the landslide literature (e.g., Dieu and Gjermundsen, 2020). In this paper we use the term "AI" in its broader connotation, which includes also the ontological method used in this paper. See below for the description of the method and definition of ontologies.

1.3 The need for standards, ontologies, and taxonomies

Consistent, well defined vocabularies and data standards are essential in computer science applications, especially in AI. For ~~the data to have any meaning, and for multiple datasets to be combined, we need consistent vocabulary that is well-defined to~~ combine multiple datasets, vocabularies must be consistent and clearly-defined. Deep learning techniques require meanings for the inputs and the outputs (~~often specified in~~ commonly specified in data storage standards such as JPEG, or WAV), but the internal representations do not have well-defined meanings, making the models very opaque (Marcus, 2018). ~~For certain other~~ Other representations, such as logical and probabilistic representations, ~~the internal reasoning is done on~~ support internal reasoning using symbols with well-defined meanings, which lend themselves to use in explanations (Marcus and Davis, 2019).

Ontologies are “a specification of the meanings of the symbols in an information system” (Poole and Mackworth, 2017). In particular, an ontology ~~stores the vocabulary used to define entities and relationships~~ defines the vocabulary for individuals and relationships within a knowledge domain. Individuals may be concrete entities (e.g. a rock), or abstract concepts, (e.g. numbers); relationships are properties that describe how individuals are connected. Typical examples of relationships include: is-a-kind-of, and is-part-of, is-superclass-of, has-some-property; the ontology also defines axioms controlling the use of the ~~voabulary~~ vocabulary for logical and thematic consistency (Poole and Mackworth, 2017). Given these axioms, the vocabulary ~~is unambiguously interpreted, and the~~ can be unambiguously interpreted according to the rules of symbolic logic, and implicit relationships between ~~the entities~~ entities or instances of those entities can be inferred(?). ~~Vocabularies should be, whenever possible, Aristotelean taxonomies. Aristotelean-~~

Vocabularies can be Aristotelian taxonomies, which are logically-consistent and multi-hierarchical. Aristotelian taxonomies are constructed by defining concepts from their relation to a more general parent concept (genus) and using differentiating properties (differentia) to distinguish concepts within the same genus (Aristotle, 350BC). For example, "Slides in soil" and "Slides in rock" share the same parent concept "Slides" and they are differentiated by the property dealing with the material type, "Soil" and "Rock", which make them uniquely identifiable. Taxonomies based on Aristotelian definitions ~~tend to be~~ support multi-hierarchical knowledge networks and can be used by computers to make logical inferences (~~?Smith, 2003~~) (Poole et al., 2009; Smith, 2003). The term ‘multi-hierarchical’ implies that there is more than one way to move through a taxonomy to arrive at a particular node or term. For example, the landslide taxonomy can be arranged based on different properties. If the landslide types are firstly arranged based on the type of movement and then based on the type of material, one path within the taxonomy would be: Landslide> slides> slides in rock and slides in soil. Alternatively, if the landslide types are arranged first based on the material type and then on the movement type, the path of the taxonomy would be: Landslide> landslides in rock> slides in rock and flows in rock. Both paths are valid, but they reach the same concept in different ways. The Natural Hazard Classification code list extension for landslides presented in this paper was prepared using the open access Aristotelian Class Editor (ACE) software (Minerva Intelligence, 2019d). Knowledge stored in a domain-specific ontology (e.g. geohazards) can be accessed by computers, allowing for data investigation through various artificial intelligence (AI) techniques, including probabilistic matching ~~as for~~ between semantic networks, the technique used in this study.

85 Significant progress has been made in the development of taxonomies for geoscience information interchange by the IUGS
CGI Geoscience Terminology Working Group which produced the GeoSciML standard along with the OGC (CGI, 2003).
However, ontology applications in Earth Sciences are scarce. Notable exceptions are in economic geology (Smyth et al., 2007),
geohazards (Jackson Jr et al., 2008), and disaster risk reduction domains (Phengsuwan et al., 2019; Sermet and Demir, 2019).
The INSPIRE framework, through its standardised vocabularies (~~“Code Lists”~~code lists), provides ~~the a~~ necessary foundation
90 upon which AI applications with explainable output can be constructed. INSPIRE application examples in landslide studies
include the LAND-deFeND Italian landslide database structure (Napolitano et al., 2018) and a deep learning algorithm to
map landslide susceptibility (~~Hajimoradlou et al., 2019~~). ~~In this~~ (Hajimoradlou et al., 2020). In Hajimoradlou et al. (2020)’s
implementation of deep learning, training features were labelled with INSPIRE-compliant semantics to enable reproducibility
of the experiment by other researchers.

95 In this study, we present an AI-based landslide susceptibility application using a natural hazard ontology. We do so by
building from the ontology created by Jackson Jr et al. (2008), and by embedding INSPIRE code lists wherever possible and
by ~~aligning input and output~~ aligning input and output data to the INSPIRE data standards.

1.4 Landslide susceptibility and hazard

Landslide susceptibility is defined as the relative spatial probability of occurrence for a landslide based on the intrinsic prop-
100 erties of a site (SafeLand, 2011). The concept of susceptibility differs from hazard in that the temporal probability of occur-
rence, the triggering factors, and the magnitude of the event are not considered in the definition of a susceptibility map (Safe-
Land, 2011; Van Den Eeckhaut and Hervás, 2012). ~~Statistical methods, physical methods~~ To produce landslide susceptibility
maps, three approaches are usually applied: statistical, physical, and expert-based ~~methods can be applied to produce landslide~~
~~susceptibility maps~~ (SafeLand, 2011). Statistical methods rely on the analysis of landslide databases and their relation to land-
105 scape properties (see review by Reichenbach et al., 2018); physical methods calculate the limit equilibrium between ~~failure~~
~~resisting forces and driving~~ failure-resisting and -driving forces in slopes (e.g., Baum et al., 2008); and expert-based methods
rely on expert opinion and the assumption that influencing factors are known and are specified in the models (Dai et al., 2002).
The AI approach used in this study ~~mimies the domain expert reasoning, providing qualitative landslide~~ is an example of
the expert-based approach, as the models follow rules that represent the reasoning process of a landslide-expert, providing
110 semi-quantitative susceptibility maps.

2 Methods

Figure 1 outlines the methodological workflow followed in this study to produce explainable landslide susceptibility assessments
in the Veneto Region of Italy. We extended INSPIRE (Section 2.1), we constructed an ontology (Section 2.2) and we defined
expert-models (Section 2.2.1) and instances, represented by mapping polygons (Section 2.2.2). We then compared the similarity
115 of models and instances to produce a matching score, which is used as susceptibility indicator (Section 2.2.3). Finally, the
results are delivered in an interactive webmap (Section 2.2.4).

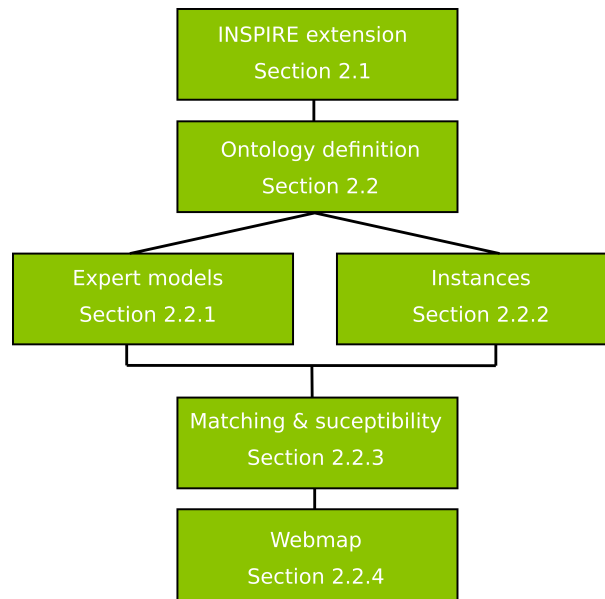


Figure 1. The workflow followed in this study and corresponding method sections. We extended INSPIRE, defined an ontology, expert models and mapping instances. We compared models and instances to deliver a susceptibility map which is available online

2.1 INSPIRE extension

Technical guideline documents outline the data structure for each theme within the INSPIRE directive, its encoding rules, its metadata standards, and some of its use cases. Data structures are formalized as formally represented using Unified Modeling Language (UML) and Extensible Markup Language (XML) application schemas. These application schemas model spatial objects, modeling thematic entities as feature types (vector-based spatial data), describe properties of, defining properties for each feature type (often with standardised vocabularies) and characterize the relationships between different, and characterizing relationships between feature types. As such, each one of these data structures Where applicable, standardised vocabularies are adopted for property value ranges. INSPIRE themes can be understood as an ontology (See Section 2.2 below), by defining various entities and the relationships between them.

INSPIRE data is generally encoded models are implemented as Geography Markup Language (GML) and can also be application schemas and serialized using Extensible Markup Language (XML). This enables data distribution provided as Open Geospatial Consortium (OGC) - compliant web services. Feature-type Geospatial features are located using vector-based spatial data. Feature properties have value types (e.g. geometry for vector data sets) and; properties whose value ranges are controlled vocabularies have values implemented as code lists. Code lists store the terminology that can be used in each domain to specify the value attribute, and may incorporate vocabularies developed outside of INSPIRE (e.g. IUGS CGI rock type taxonomy, Natura 2000 And Emerald Bio-geographical Region Classification). Some code lists within INSPIRE are not extensible, some are extensible with narrower values, and some allow additional values at any level. Code list values, definitions

and hierarchical structures are stored in the ~~INSPIRE-registry~~INSPIRE registry, making them accessible to and reusable by anyone. INSPIRE schemas can also be extended to include additional concepts and/or feature types. For this project, we worked with four INSPIRE themes: Geology, Land Cover, Hydrography and Natural Risk Zones. The Natural Risk Zone application schema was not fully adequate for this application as it lacked the ‘landslide susceptibility’ concept and ‘landslide type’ code lists (Tomas et al., 2015). We addressed this issue by formally extending the Natural Risk Zone schema and the Natural Hazards code list.

140 2.2 ~~Ontological-grounded~~Ontologically-grounded probabilistic matching

The method used to produce INSPIRE-based landslide susceptibility maps, ~~is a probabilistic comparison system that mimics human expert reasoning~~uses qualitative probabilistic reasoning that incorporates expert knowledge, making qualitative predictions based on comparisons between models and instances (~~e.g., Sharma et al., 2010; Smyth et al., 2007; ?; Smyth and Poole, 2004~~)~~–Models are expert-based~~(e.g., Sharma et al., 2010; Smyth et al., 2007; Poole and Smyth, 2005; Smyth and Poole, 2004). A
145 model is a set of rules defined a priori by an expert, based on scientific literature, making use of the entities and properties defined in the ontology. These models aim to represent expert conceptualized descriptions of a given phenomenon or entity (e.g. landslide susceptibility). The properties used in a model description are concepts stored in the ontology~~(e.g. soil slide – has slope – steep)~~, along with frequency terms (e.g. soil slide – has slope – moderately steep – always). Frequency terms used in this study ~~included~~are: “always”, “usually”, “sometimes”, “rarely” and “never”. These terms were chosen as they express
150 experience-based judgements that geoscience practitioners may use in field assessments. The term “never” allows the system to explicitly deal with negation (e.g. soil slide - has surficial material - bedrock - never). The properties and the frequency terms are encoded in semantic triple format (W3C Working Group, 2014) and the resulting model is a semantic network. Semantic networks are a graph representation of knowledge where nodes are concepts and edges are the semantic relation between concepts (Shapiro, 1992); see Figure 2 for example. Real-world areas on the ground (map units – more generally ~~referred~~
155 referred to as “instances”) are also described by semantic networks using the same properties stored in the ontology, but ~~they~~triples are accompanied by ~~true-or-false-qualifiers~~Boolean qualifiers to represent presence or absence of a specific property (e.g. polygon – has slope – steep – truepresent). Comparisons, referred to as matches, between instances and models is possible because models and instances all use the same structured terminology, as controlled by the ontology.

Similarity scores are awarded based on the type of match between instance and model properties, the semantic distance in
160 the taxonomy of compared property values and the model property frequency term (Figure 2). Match types include, ~~Exact, A Kind-exact, a kind~~ Of (AKO) exact, and ~~Possible–An-Exact-possible.~~ An exact match indicates that the property value term used in the model is present in the instance (‘a’ in Figure 2), in which case full score is awarded for this component of the compared semantic networks. An AKO exact match indicates that the attribute-property value term found in the instance is a kind of the attribute-property value term found in the model (‘b’ in Figure 2), in which case a full score is also awarded. A
165 “Possible” match occurs when the concept-property value term in the instance is broader than the concept-property value term in the model, based on the defined taxonomies, in which case the score is divided by the semantic distance between the two conceptterms. For example, ‘forest’ is a more specific type of ‘forest and semi natural areas’ (‘c’ in Figure 2) and results in

the score being divided by two. The score is lower because the instance is only possibly the kind of value that the model is looking for.

- 170 In this study, an ~~Exact~~exact match or an AKO exact match of a property with frequency “always” scores ~~+10000~~+10,000, “usually” scores 9000, “sometimes” scores 1000, “rarely” scores “100” and “never” scores ~~-10000~~-10,000; unmatched attributes are awarded -10 points. These scores are an arbitrary representation of degree of surprise that uses order of magnitude numbers to distinguish qualitative measures. For an extensive review of the probabilistic comparison method, see ~~?, Sharma et al. (2010)~~Sharma et al. (2010), ~~Smyth et al. (2007)~~Smyth et al. (2007), and ~~Smyth and Poole (2004)~~Smyth and Poole (2004) Poole and Smyth (2005), Smyth et al. (2007) and
- 175 Sharma et al. (2010). This approach has been ~~successfully~~successfully applied in economic geology to generate mineral deposit exploration targets (Smyth et al., 2007), and in geohazard mapping to produce landslide susceptibility maps (Jackson Jr et al., 2008).

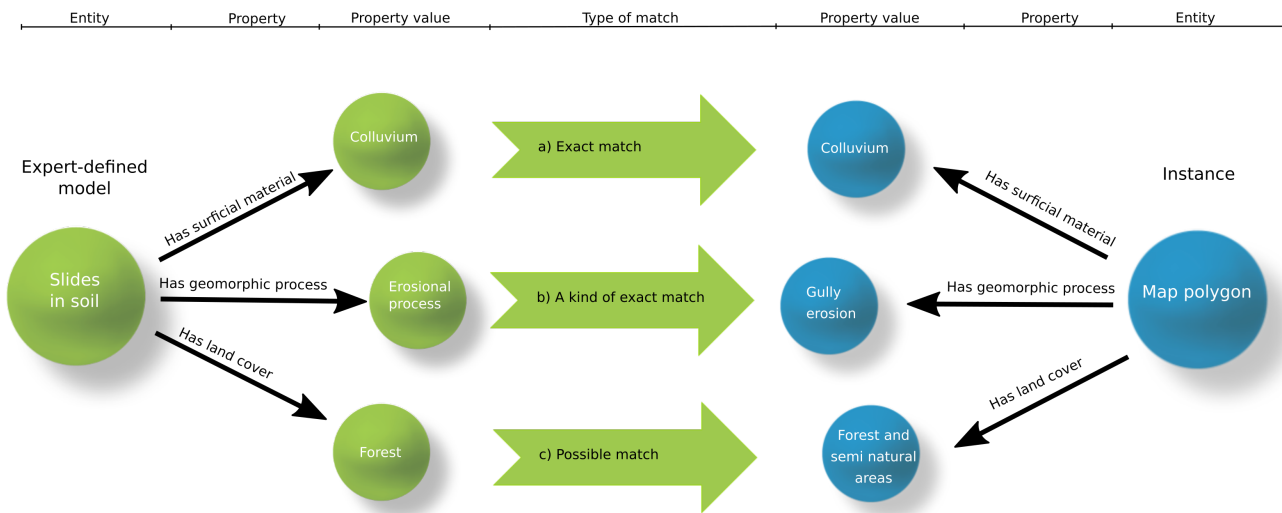


Figure 2. Graphical representation of the matching process between expert-defined models and map polygon instances. a) is an example of an Exact-exact match between the property value “Colluviumcolluvium”; b) is an example of a kind of (AKO) exact match, because “gully erosion” is a more specific kind of "erosional process". The model is looking for an "erosional process" and found a "gully erosion"; c) is an example of a possible exact match because “Forestforest and semi natural areas” is a broader concept of “Forestforest”. The model is looking for "forest" but we do not know whether the instance is a "forest". We only know that the instance is "forest and semi natural areas". The vocabulary and the hierarchy are controlled by the ontology. Note that frequency terms for model properties are not shown in this figure.

2.2.1 Landslide models

- This paper presents an AI expert-based landslide susceptibility map for three different landslide types: debris flows, slides in soil, and slides in rock (Hung et al., 2014) for the Veneto region of Italy. These three landslide types are conceptualizations of landslide models defined using knowledge recorded in the scientific literature ~~and the data available for the Veneto region.~~ These landslide models are intended to be proof-of-concept of models that can be used in the semantic approach proposed in this
- 180

paper. In particular, some of the properties used in the models are drafted from literature analysis of logging-related landslides in British Columbia, Canada (Jackson Jr, 2019). Here we briefly summarize the models, see Appendix C for a detailed explanation of each property-property value-frequency combination; detailed explanations of each property-value-frequency combination are provided in Appendix C.

185

The ‘Debris Flow’ model describes the ~~streams~~ channels that may generate a debris flow. Debris flows are flow-like landslides generated when saturated sediments move down a ~~stream~~ steep channel. They can be originated when a slide in soil intersects a flowing body of water, or when saturated bed sediments are mobilized and begin flowing downstream. Debris flows are usually triggered by intense and persistent rainfall (Hungri et al., 2014). To visualize the ‘Debris Flow’ see the table in Appendix C or navigate to <https://italy.minervageo.com/debris-flow-model/>.

190

The ‘Slides in Rock’ model describes slopes that may generate slides in rock. Slides in Rock form when steep rock slopes and cliffs fail under the influence of gravity, and are commonly triggered by intense rainfall or earthquakes. Slides in rock are usually very fast, and the failure can occur along planar, curved, and/or multiple surfaces. This model represents the collective class of landslides that have as material "rock" and movement type "slide", including rotational, planar, compound, wedge and irregular slides in rock (Hungri et al., 2014). Given the regional scale of this study, we do not have the data resolution to determine the possible failure plane geometry. For example, we cannot identify slopes more susceptible to planar rock slides rather than rotational rock slides. To visualize the ‘Slides in Rock’ model see the table in Appendix C or navigate to <https://italy.minervageo.com/the-roberti-slides-in-rock-model/>

195

The ‘Slides in Soil’ model describes slopes that may generate slides in soil. Slides in soil are downslope movements of soil under the influence of gravity, commonly triggered by intense rainfall or earthquakes. They can be slow or fast, and the failure can occur along one or many planar or curved surfaces (Hungri et al., 2014). With Slides in Soil, we refer to the collective class representing all landslides that have as material "soil" and movement type "slide", including rotational, planar, and compound, clay, silt, sand, gravel, debris slides. Given the regional scale of this study, we do not have the data resolution to determine the possible failure plane geometry and the specific kind of soil that is involved in the failure. To visualize the ‘Slides in Soil’ see the table in Appendix C or navigate to <https://italy.minervageo.com/slides-in-soil/>

205

In the presence of higher resolution information such as rock bedding orientation or shear geometry and stratigraphy in soil masses, specific kinds of rock slides (e.g. planar vs rotational) or different kinds of slides in soil (e. g. clay compound slide vs clay planar slide) susceptibility may be mapped.

2.2.2 Mapping unit and runout Map polygon instances

210

The definition of the mapping unit is a critical step in any landslide susceptibility mapping application and there are many different approaches to subdividing the area of interest to identify areas susceptible to slides in soil or rock (see review by Guzzetti et al., 1999). For this study, we used slope units, which are a geomorphic representation of single slopes bounded by drainage and divide lines (Guzzetti et al., 1999), as mapping unit ~~to identify areas susceptible to slides in soil or rock~~. We used the r.slopeunits software to ~~objectively~~ automate the slope unit delineation (~~Alvioli et al., 2016~~) (Alvioli et al., 2016, 2020). We

215

used stream line vector shapefiles provided by the Veneto Regional Government, buffered by a distance of 5 m as mapping

units to map debris flow susceptibility. In total, the region of Veneto was subdivided into 93,262 polygons, including of which 9,302 are stream buffer polygons and 83,960 are slope-unit polygons.

~~Following the calculation of susceptibility~~ We used a spatial overlay analysis to aggregate data describing the physical properties of the mapping units (Figure 6). The analysis aggregated the properties from all features that intersect the mapping units. For each property in an input layer, an aggregation type is specified as either: (a) list, whereby all of the intersecting properties are concatenated into the mapping unit (e.g. multiple rock types), or (b) Boolean evaluation, which checks whether or not the mapping unit was intersected by a specific input feature (e.g. a fault).

The properties describing each mapping unit polygon were converted into semantic networks, one network for each polygon. This conversion allows for semantic reasoning to compare and rank, based on similarity, the mapping units (hereon instances) against the expert-defined landslide models to evaluate landslide susceptibility.

2.2.3 Matching, susceptibility and runoff

The similarity score between a given model and instance is used as a proxy of landslide susceptibility. A high similarity score between an instance and a landslide susceptibility model signals a high susceptibility to that type of landslide. We deliver the similarity score between models and instances as susceptibility on the output maps.

After the susceptibility assessment, a first-order estimate of hazard is provided by calculating the likely extent of landslide runoff for the most susceptible (~~highest scoring~~ 99.9th percentile score, i. e. top one in a thousand) instances for each model. Various physical methods have been developed to calculate potential landslide runoff, given the physical properties of the material and the topography (see review by McDougall, 2016). To compute the potential runoff extents, we applied the r.avaflow code (Mergili et al., 2017) which is an open source software package implementing the two-phase debris flow model developed by Pudasaini (2012). Physical model parameters for ‘Slides in ~~Rock~~rock’ are inferred from the back-calculations of the recent Mt. Joffre landslide, in British Columbia, Canada (Friele et al., 2020), ‘Slides in ~~Soil~~soil’ and ‘Debris ~~Flow~~flow’ parameters use the default r.avaflow parameters for those landslide types (Table 1).

Various landslide size classes were simulated for each map instance, ranging from class 4 to class 6 (Jakob, 2005). Classes 4 to 6 were chosen to provide a preliminary hazard assessment, where class 4 event may have an approximate return interval of hundreds of years and class 6 are very unlikely and extreme events with return intervals on the order of thousands of years (Jakob, 2005).

2.2.4 Web map

This study’s landslide susceptibility maps and hypothetical landslide runouts for slides in soil, slides in rock and debris flows are delivered as an interactive web map based on OpenLayers (MetaCarta, 2005) and React (Facebook, 2013). Input layers are hosted through a Geoserver (The Open Planning Project, 2001) with a PostGIS (Refraction Research, 2001) backend database. INSPIRE-aligned layers are hosted on Hale Connect (WeTransform, 2014), a platform used to host and serve INSPIRE-compliant data.

Table 1. R.avaflow parameters for slides in soil, slides in rock and debris flows runoff calculations

Variables (unit)	Slides in Soil	Slides in Rock	Debris Flow
Solid fraction (%)	60	70	60
Fluid fraction (%)	40	30	40
Solid fraction internal friction angle (degree)	18	18	5
Solid fraction basal friction angle (degree)	10	10	4
Fluid fraction internal friction angle (degree)	0	0	0
Fluid fraction basal friction angle (degree)	0	0	0
Solid fraction viscosity (m ² s ⁻¹)	30	30	5
Fluid fraction viscosity (m ² s ⁻¹)	3	3	3

3 Results

3.1 INSPIRE Natural Risk Zones extension

250 To develop an INSPIRE-compliant AI application to map of landslide susceptibility, we needed to extend the INSPIRE Risk Zones theme to include the concept of landslide susceptibility and the specific code list dealing with landslide terminology. The INSPIRE extensions developed in this project are documented and stored in the [INSPIRE registry software \(Minerva Intelligence, 2019a\)](#). The Minerva [‘Minerva Re3gistry’ is \(Minerva Intelligence, 2019a\)](#), a version 1.3.1 instance of the INSPIRE registry based on the Re3gistry software (ISA, 2016). The registry service is packaged within a collection of
255 Docker (Hykes, 2013) containers and hosted on a local server.

The Natural Risk Zone core (NZ-core) schema extension, which includes the Natural Risk Zone Susceptibility feature type was based on SafeLand recommendations (SafeLand, 2011). The [‘Natural Hazard Classification’ code list](#) was extended ([Minerva Intelligence, 2019b](#)) to include a classification of various landslide types using the [Updated Varnes Landslide Classification updated Varnes landslide classification](#) (Hung et al., 2014), which [has been is a landslide classification](#) widely
260 adopted within the scientific community, [the Natural Hazard Classification Landslide Extension’ code list \(Minerva Intelligence, 2019b\)](#), and a new code list of landslide size classes ([Minerva Intelligence, 2019c](#)) based on Jakob (2005) ([see the ‘Landslide Size Class’ code list \(Minerva Intelligence, 2019e\)](#)). This [. The landslide size](#) code list contains ten landslide size classes based on landslide volume and descriptions of approximate damage potential.

3.1.1 Code list extension

265 The Natural Hazard Classification code list extension for landslides [was prepared using the open access ACE taxonomy editing software \(Minerva Intelligence, 2019d\)](#). The [classification tree](#) considers material type and failure movement, [splitting the tree, first on type of movement, and then on type of material, following Hung et al. \(2014\)](#) (Figure 3). [Properties dealing with Other properties, such as:](#) water content, depth of failure, rate of movement, loading state, channelized state, and fail-

ure plane geometry (see Appendix B) are used to ~~further describe the different landslide types. The~~ describe the individual
270 landslide types, as the unique combination of these properties allows for unambiguous classification ~~of the different landslide~~
~~types into an Aristotelean taxonomy which is multi-hierarchical and amenable to reasoning by both humans and computers.~~
~~Multi-hierarchical means that, for example, the concept “debris flow” is in different position in the taxonomy depending on the~~
~~chosen top concept. It can be three level down from the top concept “flow-like landslide” (flow landslides>flows in soil>debris~~
~~flow) or two level down from the top concept “fast landslides” (fast landslide >debris flow)in an Aristotelian taxonomy. We~~
275 used these properties because, even if not shown in the final taxonomic tree, they are explicitly applied in the wordy description
of landslide types by Hungr et al. (2014).

The formal extension registration process via the INSPIRE Registry software does not enable the representation of such multi-hierarchical classifications. Because of this we had to work with a single tree hierarchy, and consequently chose to first divide the classes on type of failure followed by a division based on the type of movement (Figure 3).

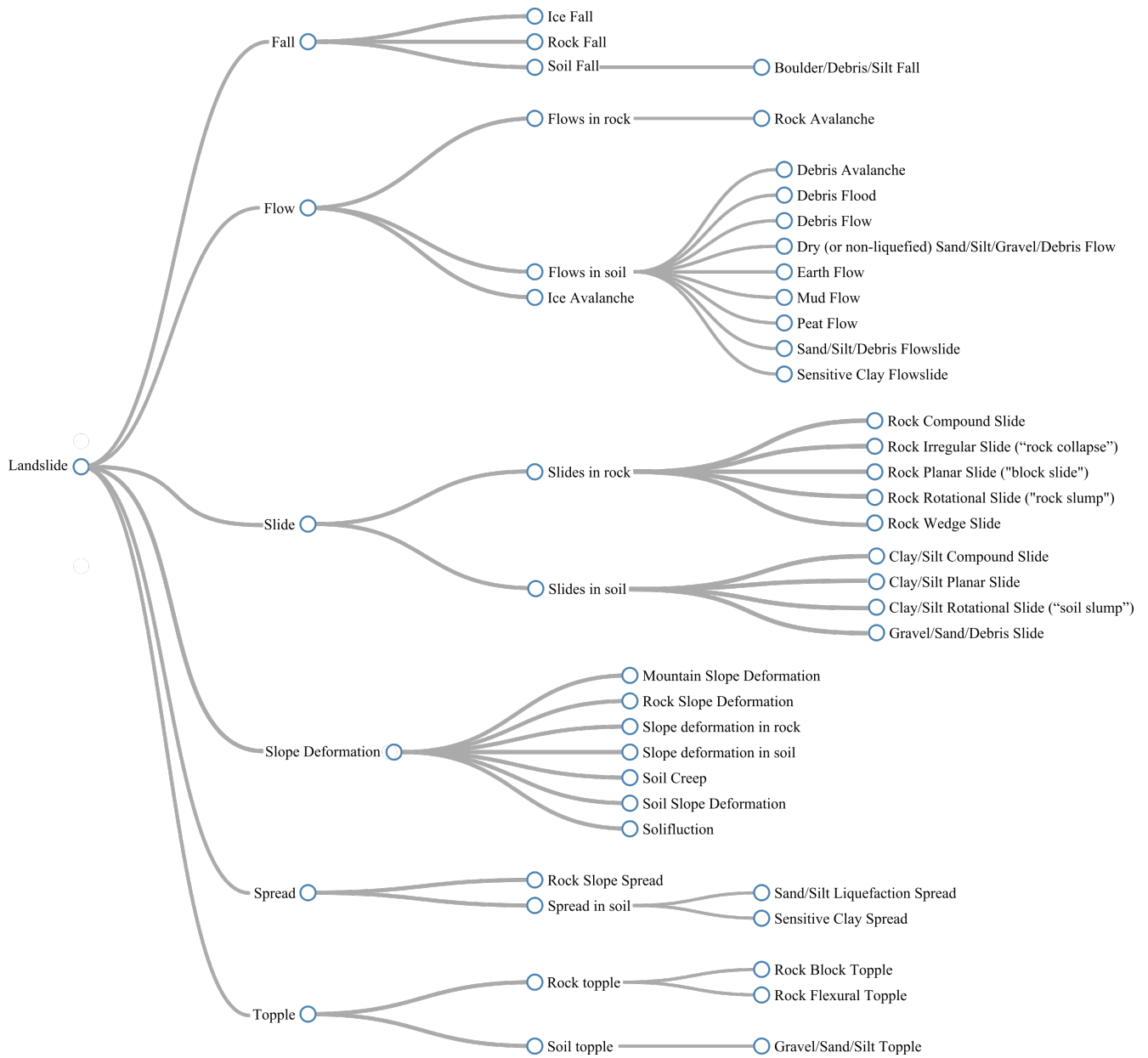


Figure 3. Natural Hazard Category code list extension for landslides

280 3.1.2 Schema extension: susceptibility

The INSPIRE Natural Risk Zone schema includes hazard and risk feature types, but the concept of susceptibility as a feature type is missing. To overcome this problem, we extended the INSPIRE Natural Risk Zone core XML schema, adding a Natural Risk Zone Susceptibility schema (Minerva Intelligence, 2019e). The Natural Risk Zone Susceptibility schema includes Abstract Susceptibility Area (a in Figure 3) and Susceptibility Area feature types (b in Figure 3 Figure 4). The Susceptibility Area feature type is modelled following the structure of the Hazard Area and Risk Zone feature types in the INSPIRE Natural Risk Zone core schema. Susceptibility Area has three elements: Geometry, Influencing Factor and Relative Spatial Likelihood of Occurrence (b in Figure 3 Figure 4). Geometry, as with all INSPIRE vector datasets, is the geometric representation of the extent of the feature on the Earth Surface as a spatial feature. Influencing factors are defined as the intrinsic, preparatory variables which make an area susceptible to a hazard (SafeLand, 2011). Influencing factors are unbounded in multiplicity (i.e. can be
290 as many as needed) and can be defined qualitatively or quantitatively. Qualitative influencing factors are expressed as a string, while quantitative influencing factors are expressed as GML:MeasureType (e in Figure 3 Figure 4). Whether defined quantitatively or qualitatively, the influencing factor can also define a DataSetType attribute, such as slope or air quality. Influencing factors are used in the calculation of Relative Spatial Likelihood of Occurrence, which is an element that can be quantitatively or qualitatively defined (d in Figure 3 Figure 4). The relative spatial likelihood of occurrence refers to values that represent the
295 spatial probability of occurrence of a specific hazard type, given the influencing factors present in the area (SafeLand, 2011). The Influencing Factor element allows end users of Susceptibility Area datasets to understand which known conditions of the specific area led to the resultant ~~Relative Spatial Likelihood of Occurrence (susceptibility)~~ susceptibility.

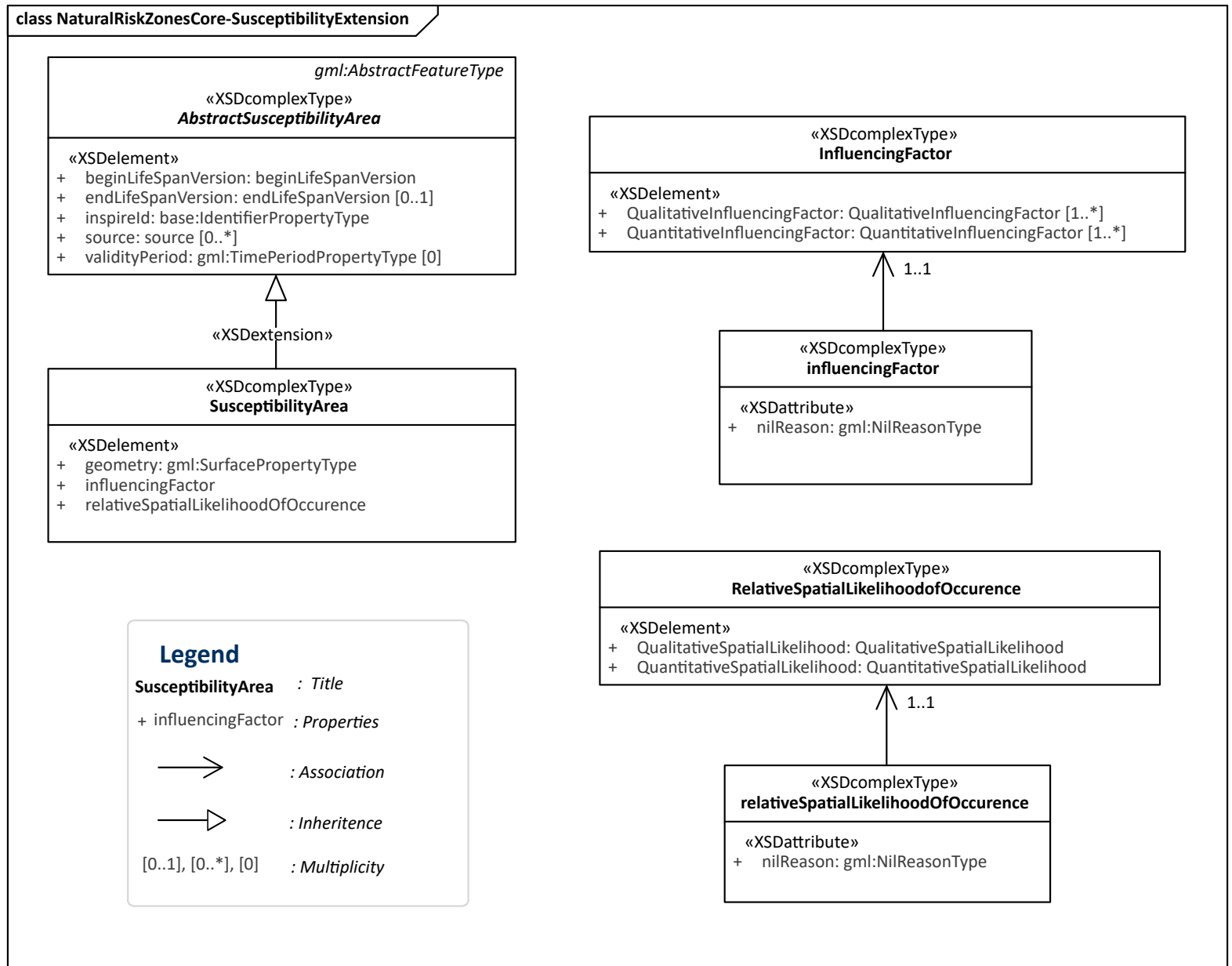


Figure 4. UML diagram showing the Natural Risk Zone Core Schema extended with Natural Risk Zone Susceptibility schema (red squares); a) Abstract susceptibility area; b) Susceptibility area; c) Influencing factor; d) Relative spatial likelihood of occurrence. [UML diagram showing Natural Risk Zone Susceptibility schema extension of the Natural Risk Zone Core Schema](#)

3.2 Landslide susceptibility mapping in Veneto

3.2.1 Input data

300 For this study, we used open access datasets from the Veneto Region Geoportal and other sources (Table 2 and 3). We Aligning all input datasets was beyond the scope of this project. We did, however, want to show the value of INSPIRE-aligned data and therefore aligned stream network, CORINE land cover, bedrock geology, and the Italian Landslide Inventory (IFFI) (Table 2) to INSPIRE ~~standards using the software program using~~ Hale Studio (WeTransform, 2008). Figure 5 shows how different tools in Hale Studio are used to align properties from the source dataset to the target dataset. For example, the field “eta” –“Age” in
305 Italian, of the original Veneto dataset, was directly mapped to four different INSPIRE fields: the olderNamedAge.href and title and the youngerNamedAge.href and title. Note that olderNamedAge.href youngerNamedAge.href are hyperlinks to the code list value id and the title is the actual code list term from the GeochronologicEraValue code list. This alignment is done with many classification methods, including: Groovy Scripts, formatted strings and assign-alignment tools. For further explanation on term alignments, refer to the documentation of Hale Studio (WeTransform, 2008). Datasets used that were not compliant
310 with INSPIRE include: lakes, watersheds, permafrost, fire, slope angle, faults, soil, roads and railways (Table 3).

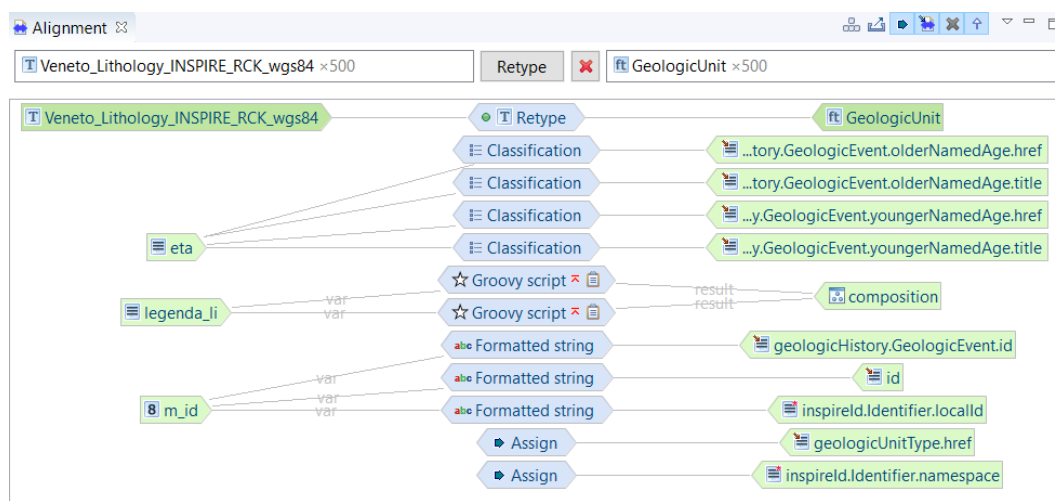


Figure 5. INSPIRE alignment visualization within Hale Studio showing the alignment from the Veneto bedrock geology Veneto Lithology "INSPIRE RCK wgs84" shapefile source (left side) to the target "Geologic Unit" feature type within the INSPIRE Geology Schema (right side).

Table 2. INSPIRE-compliant layers

Layer	Description	Source URL (last access: January 2020)
Streams	Hydrographic network in the Veneto region, including streams, rivers, and other inland flowing water bodies	https://idt2.regione.veneto.it
Land Cover (CORINE)	Land cover units in the Veneto region. The CORINE Land Cover (CLC) classification was used which includes 44 classes, and was last updated in 2018	https://land.copernicus.eu/pan-european/corine-land-cover
Geology	Bedrock lithology in the Veneto region.	http://www.pcn.minambiente.it/mattm/en/wfs-service/
IFFI Landslide Points and Areas	Landslides that have been identified in the Veneto region as part of the IFFI project. The INSPIRE Natural Hazard Category code list was extended to include the updated Varnes landslide classification (Hungar et al., 2014), and the data were aligned to this standard	http://www.pcn.minambiente.it/mattm/en/wfs-service/

Table 3. Layers not compliant with INSPIRE standards

Layer	Description	Source URL (last access: January 2020)
Lakes	Lakes in the Veneto region.	https://idt2.regione.veneto.it
Watersheds	Watersheds in the Veneto region, derived from a digital elevation model from the TINITALY project made available by the National Institute of Geophysics and Volcanology (INGV).	http://tinality.pi.ingv.it/
Permafrost	Permafrost derived from the Global Permafrost Zonation Index Map (Gruber, 2012)	http://www.geo.uzh.ch/microsite/cryodata/
Fires	Location and date of past forest fires in the Veneto region.	https://idt2.regione.veneto.it
Slope	The gradient of the slope in the Veneto region, derived from a digital elevation model from the TINITALY project made available by the National Institute of Geophysics and Volcanology.	http://tinality.pi.ingv.it/
Faults	Faults in the Veneto region, published as part of the Database of Individual Seismogenic Sources (DISS) provided by the National Institute of Geophysics and Volcanology (INGV).	http://diss.rm.ingv.it/diss/index.php/DISS321
Soils	Soil map of the Veneto region, including information about surficial deposit genesis, material, texture, thickness, geomorphic form and process.	https://idt2.regione.veneto.it
Railroads	Railroad network in the Veneto region.	https://idt2.regione.veneto.it
Roads	Road network in the Veneto region.	https://idt2.regione.veneto.it

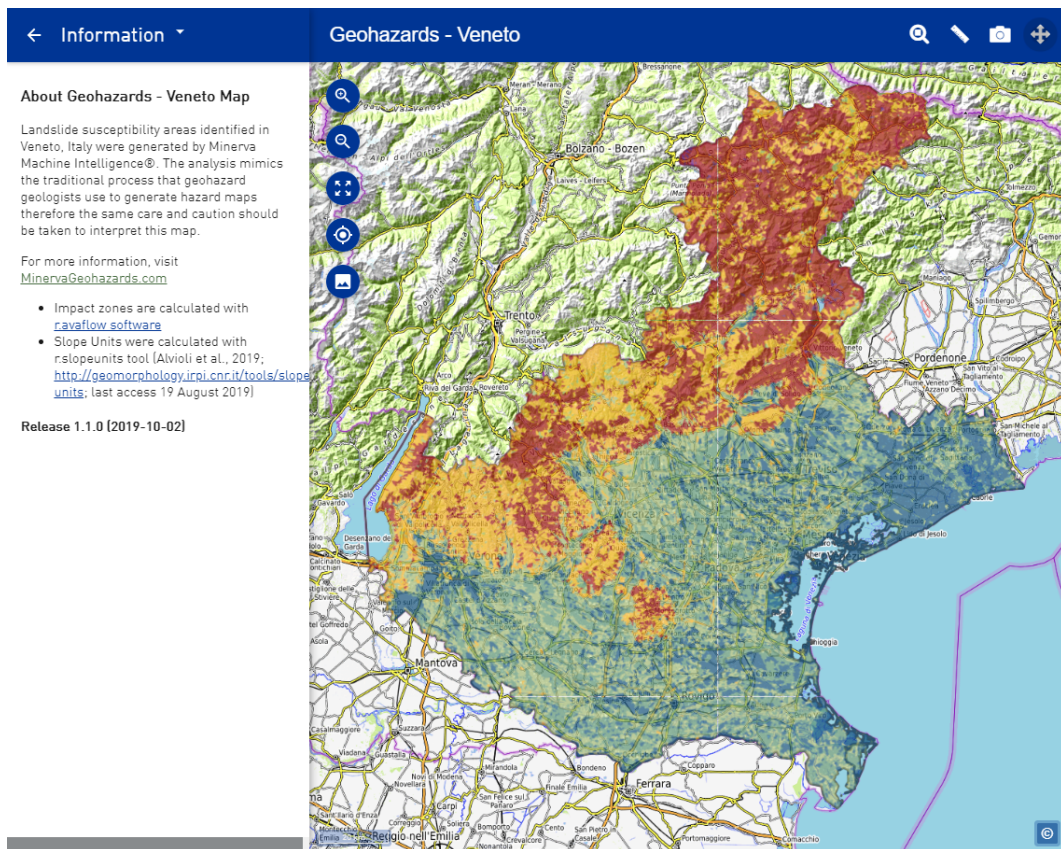


Figure 6. [Web map interface showing susceptibility to slides in soil in Veneto, Italy.](#) Colours range from blue (0-20 score percentile), to dark green (20-40 score percentile), light green (40-60 score percentile), yellow (60-80 score percentile), red (80-99.9 score percentile) and purple (99.9-100 score percentile). Base map credit: © OpenTopoMap (CC-BY-SA)

3.2.2 Mapping units and Spatial overlay Webmap

We used a spatial overlay analysis to aggregate data describing the physical properties of the mapping units,

The 83,960 slope units and 9,302 stream buffer polygons instances (Figure 6) .The analysis was conducted using a custom QGIS script which aggregated the properties from all features that intersect the mapping units. For each property in an input layer, an aggregation type is specified as either: (a)list, whereby all of the intersecting properties are concatenated into the mapping unit (e. g. multiple rock types), or (b)boolean evaluation, which checks whether or not the mapping unit was intersected by a specific input feature are encoded with the available data, then transformed from vector files into semantic network format. Then, each polygon was matched against the expert-based slides in soil, slides in rock and debris flow models and colour coded on matching score percentile to portray landslide susceptibility (Figure 6). The left-side panel of the webmap shows the landslide model layers, the reference layers, and different base maps (Figure 7). By clicking on a polygon (instance),

a popup window opens (Figure 7): this window contains the name and hyperlink to the INSPIRE registry code list definition of the landslide type investigated, the attributes that are present in the mapping unit (e.g. a fault). The end results are polygonal representations of the landscape which are attributed with all available data for landslide susceptibility mapping as required by our ontology.

325 Web map interface portraying susceptibility to slides in soil in Veneto, Italy. Base map credit: © OpenTopoMap (CC-BY-SA)

3.2.3 Semantic network conversion, matching and impact zone modelling

The properties describing each mapping unit polygon were converted into semantic networks, one network for each polygon. This conversion allows for semantic reasoning to compare and rank, based on similarity, the mapping units against the expert-defined landslide models to evaluate landslide susceptibility. We deliver the similarity score between models and instances on the output maps. A higher similarity score between an individual mapping unit and a landslide susceptibility model signals a higher susceptibility to that type of landslide over other polygons. Explanations on how the scores were calculated are shown on bedrock lithology, erosional process, etc.), the instance percentile rank and total match score, the hyperlink to the comparison of the instance against other landslide models (e.g the slides in rock model), and (only for the 99.9th percentile score, top one in one thousand) buttons to turn on the display of landslide runout for different landslide classes, and the hyperlink to the match report available for each spatial unit (Figure ??).

The match report shows matching properties with green lines and unmatched properties in beige. Conflicting properties, if present, are coloured in red. Within the match report, there are hyperlinks to explanations describing how points are awarded for each property. is a detailed table showing the results from the model-instance semantic matching, ensuring the explainability of the results. Each line corresponds to a property-value-frequency term (e.g has slope – moderately steep – always) comparison between the model and the instance, how they match (with a hyperlink to textual explanation on how the score was awarded), the numerical score value, (see Table 4 for example) a textual explanation on why that attribute was chosen, and the original data value (Table 5). An "advice" button opening a textual advice expressing which of the instance unmatched attributes may change the score is available. This advice is a sort of data-advice: it invites the user to check in the field or in some other databases if, for example, a fault is present in that specific instance.

Sample Match Report showing polygon 117309 compared to slides in soil model

Landslide runout simulations were computed for all mapping units that fell within or above the 99.9th percentile of evaluated instances for each landslide type (Figure 7). Various landslide size classes were simulated for each instance, ranging from class 4 to class 6 (Jakob, 2005) for a total of 3696 landslide runout simulations. Classes 4 to 6 were chosen to provide a preliminary hazard assessment, where class 4 event may have an approximate return interval of hundreds of years and class 6 are very unlikely and extreme events with return intervals on the order of thousands of years (Jakob, 2005).

Table 4. [Simplified match report table showing instance 117309 compared to slides in soil mode. The match report is accessible online by clicking https://spot.italy.minervageohazards.com/match_results?if_id=34434&t_id=117309](https://spot.italy.minervageohazards.com/match_results?if_id=34434&t_id=117309)

<u>Model</u>			<u>Instance</u>			<u>Results</u>	
<u>Property</u>	<u>value</u>	<u>freq.</u>	<u>Property</u>	<u>value</u>	<u>freq.</u>	<u>match type</u>	<u>Score</u>
<u>has_Geomorph Process</u>	<u>ErosionalProcess</u>	<u>always</u>	<u>has_Geomorph Process</u>	<u>Gully Erosion</u>	<u>present</u>	<u>AKO match</u>	<u>10000</u>
<u>has_Surficial Material</u>	<u>Colluvium</u>	<u>always</u>	<u>has_Surficial Material</u>	<u>Colluvium</u>	<u>present</u>	<u>exact match</u>	<u>10000</u>

Table 5. [Simplified match report table showing the comment for the model property "has erosional process" matching the instance property "Gully erosion". The full match report is accessible online at https://spot.italy.minervageohazards.com/match_results?if_id=34434&t_id=117309](https://spot.italy.minervageohazards.com/match_results?if_id=34434&t_id=117309)

<u>Model</u>	<u>Instnce</u>	<u>Comment</u>	<u>Original value</u>
<u>Erosional process - Always</u>	<u>Gully erosion - Present</u>	<u>Active erosional processes are possible indicators of landslide activity, as landslides occur where landslides have occurred before.</u>	<u>Rock fall, gully erosion, erosional process, karst.</u>

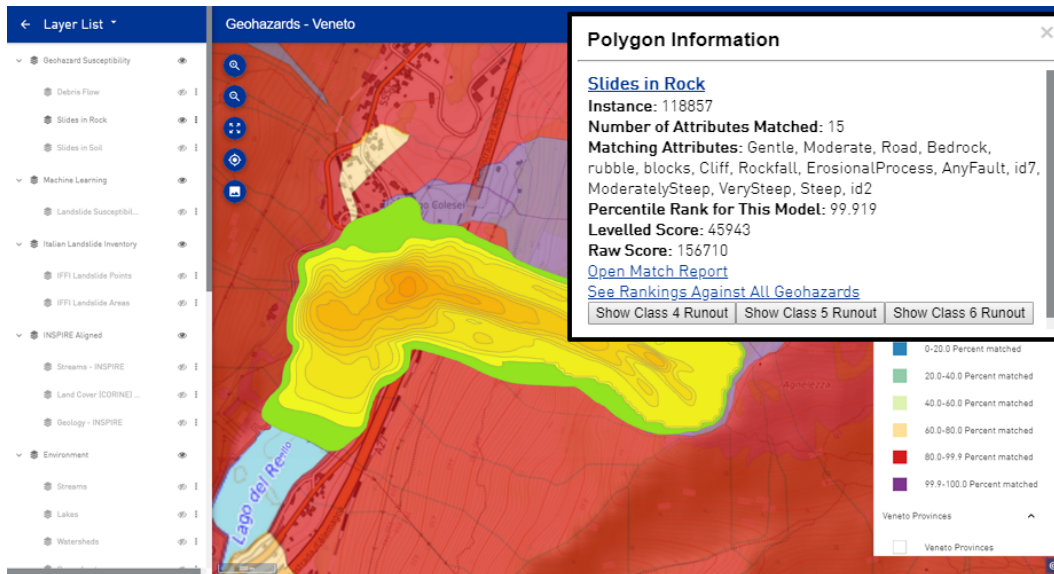


Figure 7. Landslide runoff and Screen capture showing left side panel with the layer list, information popup in Web Map window, map legend, and the landslide runoff. Base map credit: © OpenTopoMap (CC-BY-SA)

4 Discussion

4.1 INSPIRE as a framework for Explainable AI

Across society, the use of numerous complex and non-standardized earth science taxonomies results in interoperability ~~problems~~limitations,
355 which hinder the widespread implementation of explainable AI solutions to natural hazard-related problems. This is evident in
the landslide domain, where data layers for landslide susceptibility analysis, ranging from landslide databases (Van Den Eeck-
haut et al., 2013) to geomorphology maps, vary across regions and countries. Consequently, despite the wealth of scientific
literature on landslides in general, and landslide susceptibility in particular (Reichenbach et al., 2018), broad-scale operational
landslide hazard management systems are scarce (Guzzetti et al., 2020), resulting in significant human and economic losses
360 (Froude and Petley, 2018).

INSPIRE partially addresses this problem by providing standardised data structures for data-hosting and standard ~~terminologies~~
terminology to use within those structures. ~~As illustrated by this study, once it is~~ This study illustrates that, once INSPIRE-
compliant, European data can be subjected to ~~powerful AI~~ analytical methods that can be ~~efficiently and meaningfully applied~~
applied for practical application to multiple other equivalent INSPIRE-compliant data sets. For example, the same ~~landslide~~
365 ~~focused landslide-focused~~ ontology that uses terminology and knowledge models based on INSPIRE code lists used in this
project has been applied in South-Western British Columbia, Canada (Minerva Intelligence, 2019f).

By maintaining carefully curated standards, INSPIRE can play a critical role in AI applications that seek to be “explainable”
(Gilpin et al., 2019). Its code lists can be ~~imported into ontologies~~mapped into ontology properties, enabling machines to
make inferences of semantic and hierarchic relations based on data. The explainability in the application presented in this
370 study is provided in the form of a comprehensive match report, which can be opened via an information popup for each ~~slope~~
instance. The match report provides the user with complete access to the logic that drives the AI reasoning engine, allowing
interrogation of the results displayed on the map. By embedding explanations in a user-friendly interface, ontologically-based
AI can improve the understanding of complex geospatial problems by decision-makers, insurance companies and the general
public.

375 ~~As shown in this study, INSPIRE compliance increases interoperability of data and enables AI applications to identify new~~
~~insights from that data.~~ Public and private organisations, within and outside the European Union, can significantly enhance the
value of the data they collect and publish by using INSPIRE-compliant standards not only in natural hazard mapping but also
in other domains. ~~Quantification of this value has yet to be made, but calculations from (Craglia and Campagna, 2010) have~~
~~shown that the development of an SDI for interoperable exchange of spatial data can save between~~ A comparative study of
380 regional SDI in the context of INSPIRE implementation (Craglia and Campagna, 2010) showed that inefficient data access and
use at the European level results in economic losses in the 100-200 million Euro per annum in the Lombardia Region Million
Euro annual range. The same study, shows that the regional SDI of Lombardia, Italy, alone, allowed 3 Million Euro per year
savings to companies working in Environmental Impact Assessments (EIA), and Strategic Environmental Assessments (SEA).
Savings in the same order of magnitude can be expected by adopting INSPIRE standards in the geological hazard assessment
385 domain.

4.2 INSPIRE extension and limitations

INSPIRE-compliant datasets are still rare across European countries in general, and in Italy in particular (Cetl et al., 2017; Mijić and Bartha, 2018; Cho and Crompvoets, 2019). Consequently, ~~the authors were not able to find~~ we were unable to identify a jurisdiction in Europe with INSPIRE-compliant datasets for all the inputs necessary for this study. Therefore, instead
390 of using already-compliant data, a region optimal for demonstrating the inter-relationship between INSPIRE and explainable AI was chosen, and some of the data for that region was transformed ~~to~~ into INSPIRE compliance. In ~~so doing~~ doing so, the study provides both a case study of dealing with non-INSPIRE-compliant data, and an illustration of the rewards achievable by bringing a coherent set of data into INSPIRE compliance.

The code lists and application schemas in the INSPIRE Natural Risk Zone theme lacked the level of detail necessary for
395 this application. This is understandable, as given the broad scope of the directive, schemas lack the necessary granularity for specific applications. INSPIRE is intended to be used as an overarching umbrella under which domain-specific applications can find their place by extending it where necessary. The Natural Risk Zone theme (Tomas et al., 2015) and the extension presented in this work is an example of using this extension facility. Within the Natural Risk Zone theme, the Natural Hazard Category Value code list includes geological/hydrological hazards, including ‘flood’ and ‘landslide’, but the different subclasses of floods
400 and landslides are not specified. For this ~~landslide evaluation, and other applications seeking the status of interoperable and explainable, the kind of landslide susceptibility assessment, the~~ clear definition of landslide types, landslide size classes, and susceptibility ~~is was~~ is fundamental. For example, ~~the hazard posed by~~ a debris flow, which moves rapidly (~~tens of~~ meters per second), and an earth flow, which may move slowly (meters per year) ~~is very different. They present entirely different hazards;~~ they
405 can both destroy property but it is unlikely for an earth flow to result in fatalities while the opposite can be said of debris flows (Hungr et al., 2014). The definition of landslide sizes is also important: a size class 1 debris flow has a smaller impact area than a size class 6 event, but, by having a higher frequency, it may result in greater losses (Jakob, 2005).

From a data structure perspective, INSPIRE code lists cannot currently host multi-hierarchical taxonomies. This limits the nature of reasoning that can be brought to bear on them. We understand the technical difficulties in handling multi-hierarchical taxonomies, but hope that future versions of the Registry software will be able to handle these complex knowledge representa-
410 tions.

The INSPIRE Natural Risk Zone theme also lacks the definition of susceptibility as a concept and feature type. The term susceptibility is not implemented as a feature type because for most hazards (e. g floods and earthquakes) the concept is embedded within the concept of hazard likelihood (Tomas et al., 2015). This does not apply in the landslide domain where susceptibility and hazard are distinct concepts (e.g. Van Den Eeckhaut and Hervás, 2012). In this study, we implemented the
415 susceptibility feature type. Although we applied this feature type in the landslide domain, it will be useful for other natural hazard applications, when the spatial likelihood of hazard occurrence must be expressed separately from the general concept of hazard likelihood.

The extensibility of INSPIRE allows for domain-specific applications, like the approach presented in this paper, to fit within the INSPIRE framework. However, problems may also arise from the fact that INSPIRE is extensible. Extensibility allows

420 greater precision in terminology and schema for a specific application but this allows different public and private institutions
to implement separate, and eventually, incompatible extensions. For example, another landslide classification may be imple-
mented by another institution: this implementation may not be interoperable with the one presented in this study, but will have
the same INSPIRE compliance, leading to two conflicting standards. Much work remains at the level of thematic clusters to
425 implement as many standardized vocabularies and schemas as possible. Our extension is open and free, and we hope that other
entities will adopt it for other landslide applications.

4.3 **AI-based Ontological probabilistic matching for landslide susceptibility mapping**

The semantic AI system applied in this study aimed to replicate the reasoning with uncertainties typical of geological assess-
ments, ~~using~~ applying the terminology that geological and geotechnical professionals use in their daily practice (Smyth et al.,
2007). ~~As~~ Since they are based on expert-defined models, the landslide susceptibility maps produced in this study are compara-
430 ble to qualitative heuristic assessments (SafeLand, 2011). The choice of using a qualitative method for landslide susceptibility
assessment is in contrast with recent recommendations for the application of quantitative methods (Corominas et al., 2014).
However, in current ~~geological assessments~~ professional geological assessments and geomorphological mapping applications,
expert judgment is still widely applied (~~e.g., Association of Professional Engineers and Geoscientists of British Columbia, 2010~~
e.g., Association of Professional Engineers and Geoscientists of British Columbia, 2010; Guzzetti et al., 2012), and quantitative
435 (statistically and physically-based) methods rely on data that are not always available or of unknown quality. For example, land-
slide databases necessary for statistically-based susceptibility mapping are often incomplete, inaccurate, and geographically-
limited (Guzzetti et al., 2012). ~~Usually~~ Further, the geotechnical parameters necessary for running physical models are usually
approximated to carry out regional-scale studies (e.g., Mergili et al., 2014).

The semantic AI system applied in this study can be used in cases of data scarcity, and if coupled with numerical methods, can
440 improve the explainability of predictions. For example, by embedding the ontology concepts related to statistical parameters
(e.g. receiving operating curves, confidence intervals) or physical parameters (e. g. friction angles, viscosity), it will be possible
for the numerical outputs of quantitative methods to be explained in natural language, helping to reduce the gap between
scientists and decision-makers (Newman et al., 2017).

The main goal of this paper is not to present the semantic matching approach, but to show an example on how to modify
445 INSPIRE to make it possible to use it for landslide-specific applications. By suggesting these landslide-specific schema and
code list extensions, we set the ground for INSPIRE-compliant landslide susceptibility studies. Other organizations can build on
top of these extensions and future landslide susceptibility applications can be compared as they formally refer to the same data
structure and semantics. Note that we do not force any specific data and modeling variable selection, nor modeling approach
for landslide susceptibility/hazard/risk method. Such an effort is beyond the scope to this paper and, to some extent, already
450 addressed by the SafeLand project (e.g., SafeLand, 2011) rather, we provide the data structure and semantics to store and share
whichever method has been chosen by the modeler. For example, data selection for calculation of landslide susceptibility is
encompassed in the schema structure under “Influencing Factor” which are “unbounded in multiplicity and can be defined
qualitatively or quantitatively”, leaving broad range of possibilities to the modeler. Regarding the data quality, it is discussed in

the Natural Risk Zone schema and they refer to ISO standards (INSPIRE Thematic Working Group Natural Risk Zones, 2013)

455 . However, we recognize that specific code list (semantics) dealing with data quality and model uncertainty are missing. We hope that the INSPIRE thematic group will address this point.

5 Conclusions

This study presents ~~AI-based~~ an AI method, based on semantic network comparison, to produce landslide susceptibility maps using an ontology and standardized taxonomies within the framework provided by the INSPIRE Natural Risk Zone theme. This
460 method does not need an accurate landslides inventory to make predictions, as it uses qualitative probabilistic reasoning that incorporates expert knowledge. We produced susceptibility maps for debris flow, slides in soil and slides in rock for the province of Veneto, Italy; ~~in the framework provided by the INSPIRE Natural Risk Zone theme.~~ To produce the maps for specific landslide types, we extended the Natural Risk Zone theme to encompass both the concept of susceptibility and the different types of landslides. In particular, we registered a landslide classification extension of the Natural Hazard Category code list,
465 a landslide size class code list, and Susceptibility Area and Abstract Susceptibility Area feature types schema extensions. After defining the extension, we aligned key input layers (geology, streams, and land cover) to INSPIRE and, by using ~~a state-of-the-art~~ an ontologically-grounded probabilistic matching algorithm, we produced the landslide susceptibility layers. The processing outputs were mapped to the Natural Risk Zone Susceptibility schema extension. Then, potential impact zones of ~~potential landslides for difference landslide~~ landslides for multiple landslide-size classes were physically modelled for ~~the most susceptible polygons~~ a subset of the instances with the highest susceptibility scores. Finally, the results were ~~embedded presented~~ in a user-friendly interface, ~~and made available online~~ embedding plain language explanations on how the score was assign and advises on how to improve the matching.
470

We have demonstrated the value of INSPIRE-compliance by showing how it enhances information and knowledge interoperability, and allows for explainability in AI applications by standardized interrogation of their inputs and outputs. Ontologies
475 provide the formal structure for INSPIRE code lists to run algorithms similar to that applied here. The maps can explain the scientific results that they portray, and consequently improve the understanding of complex geospatial problems not only by domain experts but also by decision-makers and other non-specialized interested parties.

This study also illustrates that, in their current state of development, the INSPIRE standards are not sufficiently expressive to support complex landslide susceptibility mapping. We provided an example of how INSPIRE's extension capabilities may
480 be implemented to add the required expressivity. This extension framework ensures, through its Re3gistry register, that the expressivity extensions are documented and available to all interested parties for re-use. In so doing, it sets the context for the ongoing refinement of standards by the INSPIRE thematic committees.

6 Data availability

- The web application is available at: <https://map.italy.minervageohazards.com/>
- The schema extension is available at: <https://github.com/minervaintelligence/INSPIRE-NZ-Susceptibility>
- The code list extension is available at: <http://minerva.codes/registry>
- Data from the Italian National geoportal is available under “Attribution-NonCommercial-ShareAlike 3.0 Italy (CC BY-NC-SA 3.0 IT)” License, <https://creativecommons.org/licenses/by-nc-sa/3.0/it/deed.en>
- Data from the Veneto Geoportal are available under the “Italian Open Data License 2.0”, <https://www.dati.gov.it/content/italian-open-data-license-v20>
- CORINE land cover data is available under EEA standard re-use policy: re-use of content on the EEA website for commercial or non-commercial purposes is permitted free of charge, provided that the source is acknowledged (<http://www.eea.europa.eu/legal/copyright>)
- Tinality DEM is available upon request by sending an email to simone.tarquini@ingv.it with the subject of TINITALY DEM. Terms and Conditions of Use: Data is provided for research purposes only. Data is provided solely to the person named on this application form and should not be given to third parties. Third parties who might need access to the same dataset are required to fill their own application forms <http://tinality.pi.ingv.it/> Data from INGV is available under “Creative Commons Attribution-ShareAlike 4.0 International (CC BY-SA 4.0)” license <http://creativecommons.org/licenses/by-sa/4.0/>
- The permafrost data is available under “Attribution 3.0 Unported (CC BY 3.0)” licence. <http://www.geo.uzh.ch/microsite/cryodata/>.

B.1 Dictionary of Terms

Term	Description
Code list	Are vocabularies storing INSPIRE terms A dataset specifying terms for populating INSPIRE properties that require
CLC	CORINE land Cover, a classification system for land cover based on vegetation and land use
Feature type	Vector A data type representing a thematic entity in a domain of interest, typically with some geospatial location specified by vector based spatial data
IFFI	Italian Landslide Inventory
Instance	A data item that represents an individual, specific real-world entity, for this application an Instance-instance is a slope unit either a slope unit polygon or a stream buffer polygon.
Model	Expert defined conceptualization of some entity A conceptualization of the entities, properties and relationships in the domain. Three landslide models were used in this project; debris flow, slides in soil and slides in rock.
Ontology	Definition of entities and of the rules describing A formal representation of a conceptualization of the entities, properties and the relation between the entities in some domain of interest.
Semantic Network	A graph network of arcs and nodes storing data in semantic triple format that represent concepts in a domain of interest
Schema	The structure of the spatial data themes A representation of a data model, describes the structure of a data theme
Slope unit	Mapping A map unit polygon that is automatically derived from the terrain digital elevation model, based on defined by hydrologic drainage and divide lines
Taxonomy	Hierarchical classification scheme based on shared characteristics between entities
Triple	A semantic triple is a subject-object-predicate expression that asserts a fact, and it is the basic unit of a semantic network.

D Appendix B - Properties used for the landslide classification

D.1 ~~Properties used for the landslide classification~~

Property	Property definition	Property value	Property value definition
Type of movement	Landslide movement types (Hungre et al., 2014)	Fall	A fall starts with the detachment of soil or rock from a steep slope along a surface on which little or no shear displacement takes place. The material then descends largely through the air by falling, saltation or rolling (Cruden and Couture, 2011)
		Topple	A topple is the forward rotation of material about a point or axis below the centre of gravity of the displaced mass. (Cruden and Couture, 2011)
		Slide	A slide is a downslope movement occurring dominantly on surfaces of rupture or relatively thin zones of intense shear strain (Cruden and Couture, 2011)
		Spread	Spread is an extension of mass combined with a general subsidence of a upper fractured mass of material into softer underlying material. (Cruden and Couture, 2011)
		Flow	A flow is a spatially continuous movement in which surfaces of shear are short-lived, closely spaced and not usually preserved (Cruden and Couture, 2011).
Material	Landslide-forming material types (Hungre et al., 2014)	Slope deformation	Slow, sometime unmeasurable, deformation of slopes (Hungre et al., 2014)
		Ice	Glacier ice or other solid water on steep slopes (Hungre et al., 2014)
		Rock	Intrusive, volcanic, metamorphic, strong sedimentary, (carbonatic or arenaceous) and weak sedimentary (argillaceous) (Hungre et al., 2014)

Soil	Strong	Rock broken with hammer (Hungre et al., 2014)	
	Weak	Rock peeled with knife (Hungre et al., 2014)	
		Residual, colluvial, alluvial, lacustrine, marine, aeolian, glacial, volcanic, organic, random anthropogenic fills, engineered anthropogenic fills, mine tailings, and sanitary waste (Hungre et al., 2014).	
	Peat	Organic material (Hungre et al., 2014).	
	Debris	Low plasticity, unsorted and mixed material (Hungre et al., 2014).	
	Silt, sand, gravel, and boulders	Nonplastic (or very low plasticity), granular, sorted. Silt particles cannot be seen by eye. (Hungre et al., 2014).	
		Partly saturated	GW, GP, and GM unified soil classes (Hungre et al., 2014).
		saturated	SW, SP, and SM unified soil classes (Hungre et al., 2014).
		dry	ML unified soil class (Hungre et al., 2014).
	Mud	Plastic, unsorted, and close to Liquid Limit material. CL, CH, and CM unified soil classes (Hungre et al., 2014).	
Clay	Plastic, can be modeled into standard thread when moist, has dry strength. GC, SC, CL, MH, CH, OL, and OH unified soil classes (Hungre et al., 2014).		
	Sensitive	Sensitive or quick clay is a special type of clay prone to sudden strength loss upon disturbance. From a relatively stiff material in the undisturbed condition, an imposed stress can turn such clay into a liquid gel (Geertsema, 2013).	
	soft	Easily molded with fingers. Point of geologic pick easily pushed into shaft of handle. Easily penetrated several centimeters by thumb. (Hungre et al., 2014; USDA, 2012).	
	stiff	Indented by thumb with great effort. Point of geologic pick can be pushed in up to 1 centimeter. Very difficult to mold with fingers. Just penetrated with hand spade (Hungre et al., 2014; USDA, 2012).	

E Appendix C - [Landslide models](#)

490 E.1 **Landslide models**

Debris flow model [https:](#)

Table E1. [Debris flow model](https://italy.minervageo.com/debris-flow-model/) <https://italy.minervageo.com/debris-flow-model/>

<u>Instance Property-Value-Frequency</u>	<u>Model Definition Source</u>	<u>Comments</u>
<u>has surficial form -Fan(s)-always</u>	<u>(Goudie, 2014)</u>	<u>Fans are where debris flows deposit.</u>
<u>has surficial form -Terrace(s)-usually</u>	<u>(Goudie, 2014)</u>	<u>Terraces are formed by downcutting and lateral erosion of alluvial sediments by streams. Debris flows can generate terraces; hence, terraces can be indicator of debris flow activity.</u>
<u>has surficial form -Hummock(s)-always</u>	<u>(Howes and Kenk, 1997)</u>	<u>Hummocky topography may be indicator of landslide debris</u>
<u>has <u>water</u> -River/italy- minervageo.com/debris-flow-model/ Slides in soil model https://italy.minervageo.com/slides-in-soil/ Slides-Stream-always</u>	<u>(Howes and Kenk, 1997)</u>	<u>Debris flows occur periodically on established paths, usually gullies and first- or second- order streams</u>
<u>has rainfall - Extreme Rainfall- always</u>	<u>(Friele, 2012; Segoni et al., 2018)</u>	<u>Debris flows are triggered by intense rainfall (Segoni et al., 2018). Rainfall threshold for this study are derived from Friele (2012).</u>
<u>has rainfall -Severe Rainfall-usually</u>	<u>(Friele, 2012; Segoni et al., 2018)</u>	<u>Debris flows are triggered by intense rainfall (Segoni et al., 2018). Rainfall threshold for this study are derived from Friele (2012).</u>
<u>has <u>rainfall</u> -Moderate Rainfall-sometimes</u>	<u>(Friele, 2012; Segoni et al., 2018)</u>	<u>Debris flows are triggered by intense rainfall (Segoni et al., 2018). Rainfall threshold for this study are derived from Friele (2012).</u>
<u>has rainfall -Mild Rainfall-rarely</u>	<u>(Friele, 2012; Segoni et al., 2018)</u>	<u>Debris flows are triggered by intense rainfall (Segoni et al., 2018). Rainfall threshold for this study are derived from Friele (2012).</u>
<u>has <u>geomorph</u> process -ErosionalProcess-always</u>	<u>(Bovis and Jakob, 3299)</u>	<u>Streams with active erosional processes are more likely to experience debris flows than streams with less active</u>

has geomorph process (Guzzetti et al., 2012)
-MassMovement-always

Landslides are more likely to occur on slopes or valleys that have experienced landslides before

has been logged within years -5-10 years-always (Jackson Jr, 2019)

Landslides are extremely likely by 5 to 10 years after tree harvesting. Most of tree roots have died, and new trees are too small to provide anchoring effect with their roots on the slope.

has been logged within years -10-20 years-usually (Jackson Jr, 2019)

Landslides are likely by 10 to 20 years after tree harvesting as new trees are starting to provide anchoring effect with their roots on the slope.

has been logged within years -0-5 years-usually (Jackson Jr, 2019)

Landslides are likely by 0 to 5 years after tree harvesting as the trees are dead but some roots are still providing anchoring effect on the slope.

has fire within years -0-2 years-always (Jackson Jr, 2019)

Debris flows are very likely for 2 years after a wildfire. Water cannot infiltrate, runoff and erosion increase as the soil becomes water repellent and loses cohesion because of the fire heat.

has fire within years -3-5 years-usually (Jackson Jr, 2019)

Debris flows are likely between 3 to 5 years after a wildfire. The water-repellent soil horizon degrades but the roots of dead trees are starting to rot and they do not support the slope with their anchoring effect anymore.

has fire within years -5-10 years-always (Jackson Jr, 2019)

Debris flows are very likely between 5 to 10 years after a wildfire. Roots of dead trees decay, and they are not supporting the soil anymore as for the case of tree harvesting logging.

has fire within years - 10-20 years-usually (Jackson Jr, 2019)

Debris flows are likely between 10 to 20 years after a wildfire. The roots have lost anchoring effect and the new trees are still too small to support the slope.

has transport line -Road Resource-always (Jackson Jr, 2019)

Logging roads are the greatest aggravating factor for landslide activity as compared to undisturbed slopes.

has transport line -Road Resource Demographic-always (Jackson Jr, 2019)

Logging roads are the greatest aggravating factor for landslide activity as compared to undisturbed slopes.

has transport line -Road Unclassified Or Unknown-always (Jackson Jr, 2019)

The 'Road Unclassified Or Unknown' in this area of BC are mostly old inactive logging roads. This assessment has been done by visual evaluation of the data. Logging roads are the greatest aggravating factor for landslide activity as compared to undisturbed slopes.

has bed rock -volcanic igneous rock-always (Bovis and Jakob, 1999)

Quaternary volcanic rocks in BC have usually weak geotechnical properties. Basin underlain by these weak rocks are likely to experience frequent and large debris flow events.

has fire within years ->20 years-sometimes (Jackson Jr, 2019)

After 20 year since a wildfire, trees have regrown and the wildfire effects on slope stability have diminished.

has surficial material -Colluvium-Usually (Bovis and Jakob, 1999)

Debris flows are common is areas with easily erodible material.

has surficial material -Morainal Material (Till)-Always (Bovis and Jakob, 1999)

Debris flows are common is areas with easily erodible material.

has stream order -1-Always (Hungr et al., 2014)

Debris flows occur periodically on established paths, usually gullies and first- or second- order streams.

has stream order -2-Always (Hungr et al., 2014)

Debris flows occur periodically on established paths, usually gullies and first- or second- order streams.

<u>has stream order -3-rarely</u>	<u>(Hungr et al., 2014)</u>	<u>Debris flows occur periodically on established paths, usually gullies and first- or second- order streams.</u>
<u>has stream order -4-rarely</u>	<u>(Hungr et al., 2014)</u>	<u>Debris flows occur periodically on established paths, usually gullies and first- or second- order streams</u>
<u>has stream order -5-rarely</u>	<u>(Hungr et al., 2014)</u>	<u>Debris flows occur periodically on established paths, usually gullies and first- or second- order streams</u>
<u>has been logged within years ->20 years-sometimes</u>	<u>(Jackson Jr, 2019)</u>	<u>By 20 year since logging, trees have regrown and the roots are anchoring the soil again.</u>
<u>has geomorph process -Debris Flow-always</u>	<u>(Bovis and Jakob, 1999; Wilford et al., 2004)</u>	<u>Melton ratio (number that takes into account relief and area of a watershed) and watershed length allows discrimination of debris flow, debris flood, and flood prone fans.</u>
<u>has landslide type-debris flow-Always</u>	<u>(Hungr et al., 2014)</u>	<u>Debris flows occur periodically on established path. Determining the frequency of event is a non-trivial task, but the fact that someone mapped a debris flow in a specific channel, indicates the channel as prone to debris flows events.</u>
<u>has landslide type-Fall -usually</u>	<u>(Bovis and Jakob, 1999)</u>	<u>Any landslide types may accumulate debris in a channel that can be then mobilized into a debris flow</u>
<u>has landslide type-Flow-usually</u>	<u>(Bovis and Jakob, 1999)</u>	<u>Any landslide types may accumulate debris in a channel that can be then mobilized into a debris flow</u>
<u>has landslide type-Slide-usually</u>	<u>(Bovis and Jakob, 1999)</u>	<u>Any landslide types may accumulate debris in a channel that can be then mobilized into a debris flow</u>
<u>has landslide type-Spread-usually</u>	<u>(Bovis and Jakob, 1999)</u>	<u>Any landslide types may accumulate debris in a channel that can be then</u>

<u>has landslide type-Slope deformation-usually</u>	<u>(Bovis and Jakob, 1999)</u>	<u>Any landslide types may accumulate debris in a channel that can be then mobilized into a debris flow</u>
<u>has slope -Very steep-always</u>	<u>(Hungr et al., 2014)</u>	<u>Debris flows occur periodically on established paths, usually gullies and first- or second- order streams</u>
<u>has slope -Steep-always</u>	<u>(Hungr et al., 2014)</u>	<u>Debris flows occur periodically on established paths, usually gullies and first- or second- order streams</u>
<u>has slope -moderately steep-usually</u>	<u>(Hungr et al., 2014)</u>	<u>Debris flows occur periodically on established paths, usually gullies and first- or second- order streams</u>
<u>has slope -moderate-usually</u>	<u>(Hungr et al., 2014)</u>	<u>Debris flows occur periodically on established paths, usually gullies and first- or second- order streams</u>
<u>has slope -gentle-rarely</u>	<u>(Hungr et al., 2014)</u>	<u>Debris flows occur periodically on established paths, usually gullies and first- or second- order streams</u>
<u>has slope -plain-never</u>	<u>(Hungr et al., 2014)</u>	<u>Debris flows occur periodically on established paths, usually gullies and first- or second- order streams</u>
<u>has surficial form -cliff-always</u>	<u>(Howes and Kenk, 1997)</u>	<u>Cliffs indicate steep terrains where sediments may be mobilized as debris flows.</u>
<u>has surficial form -cones-always</u>	<u>(Howes and Kenk, 1997)</u>	<u>Cones store sediments that may be re-mobilized into debris flow.</u>
<u>has water -permafrost-always</u>	<u>(Hungr et al., 2014)</u>	<u>Permafrost degradation can destabilize sediments</u>
<u>has texture -blocks-always</u>	<u>(Howes and Kenk, 1997)</u>	<u>The presence of blocks can be indicator of landslide processes</u>
<u>has texture -rubble-always</u>	<u>(Howes and Kenk, 1997)</u>	<u>The presence of rubble is an indicator of landslide processes.</u>

Table E2. [Slides in rock model](https://italy.minervageo.com/the-roberti-slides-in-rock-model/) <https://italy.minervageo.com/the-roberti-slides-in-rock-model/>

<u>Instance Property-Value-Frequency</u>	<u>Model Definition Source</u>	<u>Comments</u>
<u>has geomorph process</u> <u>-GeneralPeriglacialProcesses-always</u>	<u>(Evans and Clague, 1994)</u>	<u>Landslides are common in periglacial environment, especially under changing climatic conditions.</u>
<u>has geomorph process</u> <u>-ErosionalProcess-always</u>	<u>(Guzzetti et al., 2012)</u>	<u>Active erosional processes are possible indicator of landslide activity, as landslides occur where landslides have occurred before.</u>
<u>has geomorph process</u> <u>-MassMovement-always</u>	<u>(Guzzetti et al., 2012)</u>	<u>Active mass movement processes are possible indicator of landslide activity, as landslides occur where landslides have occurred before.</u>
<u>has slope -Very Steep-always</u>	<u>(Hungr et al., 2014)</u>	<u>Very Steep slopes are prone to slides</u>
<u>has slope -Steep-always</u>	<u>(Hungr et al., 2014)</u>	<u>Steep slopes are prone to slides</u>
<u>has slope -Moderately Steep-usually</u>	<u>(Hungr et al., 2014)</u>	<u>Moderately steep slopes are prone to slides</u>
<u>has slope -Moderate-sometimes</u>	<u>(Hungr et al., 2014)</u>	<u>Moderate slopes may be prone to slides</u>
<u>has slope -Gentle-never</u>	<u>(Hungr et al., 2014)</u>	<u>Gentle slopes are rarely prone to slides</u>
<u>has slope -Plain-never</u>	<u>(Hungr et al., 2014)</u>	<u>Plain slopes are rarely prone to slides.</u>
<u>has surficial material -Bedrock-usually</u>	<u>(Hungr et al., 2014)</u>	<u>'bedrock' mapped as surficial material indicates the presence of cliffs and bluffs, possibility prone to rock slides.</u>
<u>has surficial material -Weathered Bedrock-always</u>	<u>(Hungr et al., 2014)</u>	<u>Weather bedrock is more likely to fail than fresh bedrock.</u>
<u>has weather threshold -Extreme Weather-always</u>	<u>(Friele, 2012; Segoni et al., 2018)</u>	<u>Landslides can be triggered by intense rainfall (Segoni et al., 2018) or snowmelt. Rainfall threshold for this study are derived from (Friele, 2012).</u>

<u>has weather threshold -Severe</u> <u>Weather-usually</u>	<u>(Friele, 2012; Segoni et al., 2018)</u>	<u>Landslides can be triggered by</u> <u>intense rainfall (Segoni et al., 2018) or</u> <u>snowmelt. Rainfall threshold for this</u> <u>study are derived from (Friele, 2012).</u>
<u>has weather threshold -Mild</u> <u>Weather-rarely</u>	<u>(Friele, 2012; Segoni et al., 2018)</u>	<u>Landslides can be triggered by</u> <u>intense rainfall (Segoni et al., 2018) or</u> <u>snowmelt. Rainfall threshold for this</u> <u>study are derived from (Friele, 2012).</u>
<u>has weather threshold -Moderate</u> <u>Weather-sometimes</u>	<u>(Friele, 2012; Segoni et al., 2018)</u>	<u>Landslides can be triggered by</u> <u>intense rainfall (Segoni et al., 2018) or</u> <u>snowmelt. Rainfall threshold for this</u> <u>study are derived from (Friele, 2012).</u>
<u>has land use -Alpine-always</u>	<u>(Evans and Clague, 1994)</u>	<u>Landslides are common in the Alpine</u> <u>zone, especially under changing</u> <u>climatic conditions</u>
<u>has land use</u> <u>-SubAlpineAvalancheChutes-always</u>	<u>(Hungr et al., 2014)</u>	<u>Rock slides can occur in gullies that are</u> <u>also avalanche tracks</u>
<u>has stream order -1-always</u>	<u>(Strahler, 1957)</u>	<u>Stream erosion can affect slope stability</u>
<u>has stream order -2-always</u>	<u>(Strahler, 1957)</u>	<u>Stream erosion can affect slope stability</u>
<u>has stream order -3-always</u>	<u>(Strahler, 1957)</u>	<u>Stream erosion can affect slope stability</u>
<u>has stream order -4-usually</u>	<u>(Strahler, 1957)</u>	<u>Stream erosion can affect slope stability</u>
<u>has stream order -5-sometimes</u>	<u>(Strahler, 1957)</u>	<u>Stream erosion can affect slope stability</u>
<u>has transport line -Road</u> <u>Resource-usually</u>	<u>(Jackson Jr, 2019)</u>	<u>Logging roads are the greatest</u> <u>aggravating factor for landslide</u> <u>activity as compared to undisturbed</u> <u>slopes.</u>
<u>has transport line -Road Unclassified Or</u> <u>Unknown-usually</u>	<u>(Jackson Jr, 2019)</u>	<u>Roads are an aggravating factor for</u> <u>landslide activity as compared to</u> <u>undisturbed slopes.</u>
<u>has transport line -Trail-usually</u>	<u>(Jackson Jr, 2019)</u>	<u>Roads are an aggravating factor for</u> <u>landslide activity as compared to</u> <u>undisturbed slopes.</u>

<u>has transport line -Road Recreation Demographic-sometimes</u>	<u>(Jackson Jr, 2019)</u>	<u>Roads are an aggravating factor for landslide activity as compared to undisturbed slopes.</u>
<u>has water -Permafrost-always</u>	<u>(Jackson Jr, 2019)</u>	<u>Landslides are common in periglacial environment, especially under changing climatic conditions.</u>
<u>has bed rock -metamorphic rock-always</u>	<u>(Hungur et al., 2014)</u>	<u>Foliated metamorphic rocks are usually weak and prone to failure.</u>
<u>Has CORINE land cover-Glacier and perpetual snow-always</u>	<u>(Evans and Clague, 1994)</u>	<u>Landslides are common in the Alpine zone, especially under changing climatic conditions.</u>
<u>has CORINE land cover-Bare rocks-always</u>	<u>(Hungur et al., 2014)</u>	<u>Rock outcrops can be steep and prone to landslides</u>
<u>has CORINE land cover-Road and rail networks and associated lands-always</u>	<u>(Jackson Jr, 2019)</u>	<u>Roads and rail increase landslide activity as they are a break in slope where water can accumulate</u>
<u>has fault -Any Fault-always</u>	<u>(Reichenbach et al., 2018)</u>	<u>Faults are indicator of weak rocks, and the presence of faults is one of the main parameters considered in landslide susceptibility mapping.</u>
<u>has landslide type-Rock Fall-usually</u>	<u>(Guzzetti et al., 2012)</u>	<u>Landslides are more likely to occur on slopes or valleys that have experienced landslides before.</u>

has landslide type-Rock Slope Spread-usually (Guzzetti et al., 2012)

Landslides are more likely to occur on slopes or valleys that have experienced landslides before

has landslide type-Rock topples-usually (Guzzetti et al., 2012)

Landslides are more likely to occur on slopes or valleys that have experienced landslides before

has landslide type-Slides in rock model — <https://italy.minervageo.com/the-roberti-slides-in-rock-model/>
Rock-always (Guzzetti et al., 2012)

Landslides are more likely to occur on slopes or valleys that have experienced landslides before

has landslide type-Slides in soil-sometimes (Guzzetti et al., 2012)

Note that location must also be considered. In essence, where there is soil, it is less likely that there will be steep slopes, but soil slides are a sign of an unstable slope, and therefore are not explicitly negatively correlated to rock slides

has landslide type-Slope deformation in rock-usually (Guzzetti et al., 2012)

Landslides are more likely to occur on slopes or valleys that have experienced landslides before

has landslide type-Flows in soil-sometimes (Guzzetti et al., 2012)

Where there is soil, it is less likely that there will be steep slopes, and rock slides. But soil slides are a sign of an unstable slope, and therefore are not explicitly negatively correlated to rock slides

has landslide type-Soil Fall-sometimes (Guzzetti et al., 2012)

Where there is soil, it is less likely that there will be steep slopes, and rock slides. But soil slides are a sign of an unstable slope, and therefore are not explicitly negatively correlated to rock slides

has landslide type-Slope deformation in soil-sometimes (Guzzetti et al., 2012)

Where there is soil, it is less likely that there will be steep slopes, and rock slides. But soil slides are a sign of an unstable slope, and therefore are not explicitly negatively correlated to rock slides

has landslide type-Soil Topple-sometimes (Guzzetti et al., 2012)

Where there is soil, it is less likely that there will be steep slopes, and rock slides. But soil slides are a sign of an unstable slope, and therefore are not explicitly negatively correlated to rock slides

has surficial form -cliff-always (Hungr et al., 2014)

Cliffs can generate rock slides

has texture-rubble-Always (Howes and Kenk, 1997)

The presence of blocks can be indicator of landslide processes

has texture-blocks-Always (Howes and Kenk, 1997)

The presence of rubble is an indicator of landslide processes.

has surficial form -Cones-Always (Howes and Kenk, 1997)

Cones may be formed by rock slide debris, hence they can be considered an indicator of rockslide activity

Table E3. [Slides in soil model](https://italy.minervageo.com/slides-in-soil/) <https://italy.minervageo.com/slides-in-soil/>

<u>Instance Property-Value-Frequency</u>	<u>Model Definition Source</u>	<u>Comments</u>
<u>has surficial material -Morainal Material (Till)-always</u>	<u>(Jackson Jr et al., 2008)</u>	<u>Soil slides can be generated when morainal material fails from a slope</u>
<u>has surficial material -Bedrock-sometimes</u>	<u>(Jackson Jr et al., 2008)</u>	<u>There may be some soil even when 'bedrock' has been mapped as principal surficial material</u>
<u>has surficial material -Colluvium-always</u>	<u>(Jackson Jr et al., 2008)</u>	<u>Soil slides can be generated when colluvium has been mapped as principal surficial material</u>
<u>has geomorph process -ErosionalProcess-always</u>	<u>(Guzzetti et al., 2012)</u>	<u>Active erosional processes are possible indicator of landslide activity, as landslides occur where landslides have occurred before.</u>
<u>has geomorph process -MassMovement-always</u>	<u>(Guzzetti et al., 2012)</u>	<u>Active mass movement processes are possible indicator of landslide activity, as landslides occur where landslides have occurred before.</u>
<u>has slope -Plain-rarely</u>	<u>(Hungr et al., 2014)</u>	<u>Soil slides rarely occur on plain slopes.</u>
<u>has slope -Gentle-rarely</u>	<u>(Hungr et al., 2014)</u>	<u>Soil slides rarely occur on plain slopes.</u>
<u>has slope -Moderate-usually</u>	<u>(Hungr et al., 2014)</u>	<u>Soil slides usually occur on moderate slopes.</u>
<u>has slope -Moderately Steep-usually</u>	<u>(Hungr et al., 2014)</u>	<u>Soil slides usually occur on moderate steep slopes.</u>
<u>has slope -Steep-rarely</u>	<u>(Hungr et al., 2014)</u>	<u>Soil slides rarely occur on moderate steep slopes, because usually there is not much soil on steep slopes.</u>
<u>has slope -Very Steep-never</u>	<u>(Hungr et al., 2014)</u>	<u>Soil slides rarely occur on steep slopes, because usually there is not much soil on steep slopes.</u>
<u>has land use -Alpine-never</u>	<u>(Hungr et al., 2014)</u>	<u>Soil slides rarely occur in the Alpine</u>

<u>has land use -SubAlpineAvalancheChutes-usually</u>	<u>(Hungre et al., 2014)</u>	<u>Soil slides can occur in the gullies that are also avalanche tracks.</u>
<u>has stream order -1-always</u>	<u>(Strahler, 1957)</u>	<u>Stream erosion can cause soil slides</u>
<u>has stream order -2-always</u>	<u>(Strahler, 1957)</u>	<u>Stream erosion can cause soil slides</u>
<u>has stream order -3-usually</u>	<u>(Strahler, 1957)</u>	<u>Stream erosion can cause soil slides</u>
<u>has stream order -4-usually</u>	<u>(Strahler, 1957)</u>	<u>Stream erosion can cause soil slides</u>
<u>has stream order -5-sometimes</u>	<u>(Strahler, 1957)</u>	<u>Large stream erosion may cause soil slides</u>
<u>has transport line -Trail Skid-always</u>	<u>(Jackson Jr, 2019)</u>	<u>Trail skid are aggravating factor for landslide activity as compared to undisturbed slopes</u>
<u>has transport line -Trail-sometimes</u>	<u>(Jackson Jr, 2019)</u>	<u>Trails are an aggravating factor for landslide activity as compared to undisturbed slopes</u>
<u>has transport line -Road Resource-always</u>	<u>(Jackson Jr, 2019)</u>	<u>Logging roads are the greatest aggravating factor for landslide activity as compared to undisturbed slopes.</u>
<u>has transport line -Road Unclassified Or Unknown-always</u>	<u>(Jackson Jr, 2019)</u>	<u>Roads are an aggravating factor for landslide activity as compared to undisturbed slopes.</u>
<u>has transport line -Highway-rarely</u>	<u>(Jackson Jr, 2019)</u>	<u>Roads are an aggravating factor for landslide activity as compared to undisturbed slopes.</u>
<u>has transport line -Road Recreation Demographic-sometimes</u>	<u>(Jackson Jr, 2019)</u>	<u>Roads are an aggravating factor for landslide activity as compared to undisturbed slopes.</u>
<u>has thickness -Blanket-always</u>	<u>(Jackson Jr et al., 2008)</u>	<u>Soil Slides can occur when there is enough soil that can be mobilized on a slope.</u>
<u>has thickness -Mantle of Variable</u>	<u>(Jackson Jr et al., 2008)</u>	<u>Soil Slides can occur when there is</u>

<u>has thickness -Thin Veneer-rarely</u>	<u>(Jackson Jr et al., 2008)</u>	<u>Soil Slides can occur when there is enough soil that can be mobilized on a slope.</u>
<u>has rainfall -Extreme Rainfall-always</u>	<u>(Friele, 2012; Segoni et al., 2018)</u>	<u>Landslides can be triggered by intense rainfall (Segoni et al., 2018) or snowmelt. Rainfall threshold for this study are derived from (Friele, 2012).</u>
<u>has rainfall -Severe Rainfall-usually</u>	<u>(Friele, 2012; Segoni et al., 2018)</u>	<u>Landslides can be triggered by intense rainfall (Segoni et al., 2018) or snowmelt. Rainfall threshold for this study are derived from (Friele, 2012).</u>
<u>has rainfall -Moderate Rainfall-sometimes</u>	<u>(Friele, 2012; Segoni et al., 2018)</u>	<u>Landslides can be triggered by intense rainfall (Segoni et al., 2018) or snowmelt. Rainfall threshold for this study are derived from (Friele, 2012).</u>
<u>has rainfall -Mild Rainfall-rarely</u>	<u>(Friele, 2012; Segoni et al., 2018)</u>	<u>Landslides can be triggered by intense rainfall (Segoni et al., 2018) or snowmelt. Rainfall threshold for this study are derived from (Friele, 2012).</u>
<u>has bed rock -metamorphic rock-always</u>	<u>(Bovis and Jakob, 1999)</u>	<u>Metamorphic foliated rocks have usually weak geotechnical properties. Basin underlain by these weak rocks are likely to experience more landslides compared to basin underlain by stronger lithologies.</u>
<u>has texture -blocks-always</u>	<u>(Howes and Kenk, 1997)</u>	<u>The presence of block can be indicator of mass movement processes</u>
<u>has texture -rubble-always</u>	<u>(Howes and Kenk, 1997)</u>	<u>The presence of rubble is an indicator of mass movement processes.</u>
<u>has been logged within years ->20 years-sometimes</u>	<u>(Jackson Jr, 2019)</u>	<u>By 20 year since logging, trees have regrown and the roots are anchoring the soil again</u>
<u>has been logged within years -10-20 years-usually</u>	<u>(Jackson Jr, 2019)</u>	<u>Landslides are likely by 10 to 20 years after tree harvesting as new trees are</u>

<u>has been logged within years -0-5 years-usually</u>	<u>(Jackson Jr, 2019)</u>	<u>Landslides are likely by 0 to 5 years after tree harvesting as the trees are dead but some roots are still providing anchoring effect on the slope.</u>
<u>has fire within years ->20 years-sometimes</u>	<u>(Jackson Jr, 2019)</u>	<u>After 20 year since a wildfire, trees have regrown and the wildfire effects on slope stability have diminished.</u>
<u>has fire within years -10-20 years-sometimes</u>	<u>(Jackson Jr, 2019)</u>	<u>Landslides are likely between 10 to 20 years after a wildfire. The roots have lost anchoring effect and the new trees are still too small to support the slope.</u>
<u>has fire within years -0-2 years-always</u>	<u>(Jackson Jr, 2019)</u>	<u>Landslides are very likely for 2 years after a wildfire. Water cannot infiltrate, runoff and erosion increase as the soil becomes water repellent and loses cohesion because of the fire heat</u>
<u>has fire within years -3-5 years-usually</u>	<u>(Jackson Jr, 2019)</u>	<u>Landslides are likely between 3 to 5 years after a wildfire. The water-repellent soil horizon degrades but the roots of dead trees are starting to rot and they do not support the slope with their anchoring effect anymore.</u>
<u>has fault -Any Fault-always</u>	<u>(Reichenbach et al., 2018)</u>	<u>The presence of fault is an important factor to determine landslide susceptibility</u>
<u>has fire within years -5-10 years-always</u>	<u>(Jackson Jr, 2019)</u>	<u>Landslides are very likely between 5 to 10 years after a wildfire. Roots of dead trees decay, and they are not supporting the soil anymore as for the case of tree harvesting logging.</u>

has landslide type-Slides in soil-always (Guzzetti et al., 2012)

Landslides are more likely to occur on slopes or valleys that have experienced landslides before.

has landslide type-Fall in (Guzzetti et al., 2012)
rock-sometimes

Where there is rock, it is less likely that there will be soil slides rather than landslides in rock. But landslides in rock are a sign of an unstable slope, and therefore are not explicitly negatively correlated to soil slides

has landslide type-Rock (Guzzetti et al., 2012)
topples-sometimes

Where there is rock, it is less likely that there will be soil slides rather than landslides in rock. But landslides in rock are a sign of an unstable slope, and therefore are not explicitly negatively correlated to soil slides

has landslide type-Flows in (Guzzetti et al., 2012)
rock-sometimes

Where there is rock, it is less likely that there will be soil slides rather than landslides in rock. But landslides in rock are a sign of an unstable slope, and therefore are not explicitly negatively correlated to soil slides

has landslide type-slides in (Guzzetti et al., 2012)
rock-sometimes

Where there is rock, it is less likely that there will be soil slides rather than landslides in rock. But landslides in rock are a sign of an unstable slope, and therefore are not explicitly negatively correlated to soil slides

has landslide type-Slope deformation in (Guzzetti et al., 2012)
rock -sometimes

Where there is rock, it is less likely
that there will be soil slides rather than
landslides in rock. But landslides in
rock are a sign of an unstable slope, and
therefore are not explicitly negatively
correlated to soil slides

has landslide type-Spread in (Guzzetti et al., 2012)
rock-sometimes

Where there is rock, it is less likely
that there will be soil slides rather than
landslides in rock. But landslides in
rock are a sign of an unstable slope, and
therefore are not explicitly negatively
correlated to soil slides

Author contributions.

- Gioachino Roberti, ~~Jake~~ Jakob McGregor, Clinton Smyth and David Poole wrote the paper
- 495 – Gioachino Roberti conceptually designed the susceptibility schema and landslide extension, the expert-based landslide models and expanded the geohazard ontology.
- Jakob McGregor implemented the INSPIRE schema and code list extension and designed the web map application.
- David Poole and Clinton Smyth designed the qualitative probabilistic method used to calculate susceptibility.
- Sharon Lam and Blake Boyko implemented and maintain the web map.
- Victoria Wang implemented and maintained the geohazard ontology.
- 500 – Bryan ~~and Chris~~ Barnhart and Chris Ahern implemented the qualitative probabilistic algorithm.
- ~~Steve~~ Stephen Richards supported the semantic implementations and edited the manuscript.
- David Bigelow helped in the redaction of the manuscript, reviewed the landslide models and the code list extensions.

Competing interests. The authors declare that they have no conflict of interest

Acknowledgements. This project was first presented at the Helsinki 2019 INSPIRE data challenge, ~~where~~ and won the first ~~price of the~~ data challenge prize. The authors would like to acknowledge the conference organizer committee including the National Land Survey of Finland and Ministry of Agriculture, and the Joint Research Centre of the European Commission and Spatineo. The authors would also like to acknowledge M. Alvioli et al., for the availability of the r.slopeunit code, and M. Mergili and S. Pudasaini, for the r.avaflo code, and WeTransform GMBH for the HALE Connect and HALE Studio software licences. And the reviewers I. Marchesini and O. F. Althuwaynee for constructive feedbacks on the manuscript. This research was funded by Minerva Intelligence Inc.

505

510 References

- Alvioli, M., Marchesini, I., Reichenbach, P., Rossi, M., Ardizzone, F., Fiorucci, F., and Guzzetti, F.: Automatic delineation of geomorphological slope units with r.slopeunits v1.0 and their optimization for landslide susceptibility modeling, *Geoscientific Model Development*, 9, 3975–3991, <https://doi.org/10.5194/gmd-9-3975-2016>, 2016.
- Alvioli, M., Guzzetti, F., and Marchesini, I.: Parameter-free delineation of slope units and terrain subdivision of Italy, *Geomorphology*, 358, 107–124, <https://doi.org/10.1016/j.geomorph.2020.107124>, <http://www.sciencedirect.com/science/article/pii/S0169555X20300969>, 2020.
- Aristotle: *The Categories*, 350BC.
- Association of Professional Engineers and Geoscientists of British Columbia: *Guidelines for Legislated Landslide Assessments for proposed residential development in BC*, Tech. Rep. May, 2010.
- 520 Baum, R. L., Savage, W. Z., and Jonathan W., G.: *Trigrs—A Fortran Program for Transient Rainfall Infiltration and Grid-Based Regional Slope-Stability Analysis*, Version 2.0, Tech. rep., 2008.
- Bovis, M. and Jakob, M.: The role of debris supply conditions in predicting debris flow activity, *Earth Surface Processes and Landforms*, 24, 1039–1054, [https://doi.org/10.1002/\(SICI\)1096-9837\(199910\)24,1999](https://doi.org/10.1002/(SICI)1096-9837(199910)24,1999).
- Cetl, V., Nunes De Lima, V., Tomas, R., Lutz, M., D'eukenio, J., Nagy, A., and Robbrecht, J.: Summary Report on Status of implementation of the INSPIRE Directive in EU, <https://doi.org/10.2760/162895>, http://publications.jrc.ec.europa.eu/repository/bitstream/JRC109035/jrc109035_{_}jrc109035_{_}jrc_{_}inspire_{_}eu_{_}summaryreport_{_}online.pdf, 2017.
- 525 CGI: GeoSciML, <http://www.geosciml.org/>, 2003.
- Cho, G. and Cromptvoets, J.: The INSPIRE directive: some observations on the legal framework and implementation, *Survey Review*, 51, 310–317, <https://doi.org/10.1080/00396265.2018.1454686>, <https://doi.org/10.1080/00396265.2018.1454686>, 2019.
- 530 Corominas, J., van Westen, C., Frattini, P., Cascini, L., Malet, J. P., Fotopoulou, S., Catani, F., Van Den Eeckhaut, M., Mavrouli, O., Agliardi, F., Pitilakis, K., Winter, M. G., Pastor, M., Ferlisi, S., Tofani, V., Hervás, J., and Smith, J. T.: Recommendations for the quantitative analysis of landslide risk, *Bulletin of Engineering Geology and the Environment*, 73, 209–263, <https://doi.org/10.1007/s10064-013-0538-8>, 2014.
- Craglia, M. and Campagna, M.: Advanced Regional SDI in Europe : Comparative cost-benefit evaluation and impact assessment perspectives, *International Journal of Spatial Data Infrastructures Research*, 5, 145–167, <https://doi.org/10.2902/1725-0463.2010.05.art6>, 2010.
- 535 Cruden, D. M. and Couture, R.: *The Working Classification of Landslides: material matters*, Pan-Am CGS Geotechnical Conference, p. 7, 2011.
- Dai, F. C., Lee, C. F., and Ngai, Y. Y.: Landslide risk assessment and management: An overview, *Engineering Geology*, 64, 65–87, [https://doi.org/10.1016/S0013-7952\(01\)00093-X](https://doi.org/10.1016/S0013-7952(01)00093-X), 2002.
- Dieu, T. B. and Gjermundsen, E. F.: Advance Geospatial Artificial Intelligence for Landslide Modeling, Prediction and Management, Special issue of ISPRS International Journal of Geo-Information, https://www.mdpi.com/journal/ijgi/special_{_}issues/AI_{_}landslide?listby=type{&}view=abstract, 2020.
- 540 European Commission: *INSPIRE Helsinki 2019*, <https://www.inspire-helsinki-2019.fi/>, 2019.
- European Parliament and the Council: Directive 2007/2/EC of the European Parliament and of the council of 14 March 2007 establishing an Infrastructure for Spatial Information in the European Community (INSPIRE), *Official Journal of the European Union*, 50, 1–14, <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2007:108:0001:0014:EN:PDF>, 2007.
- 545

- European Parliament and the Council: Commission Regulation (EU) No 1312/2014 of 10 December 2014 amending Regulation (EU) No 1089/2010 implementing Directive 2007/2/EC of the European Parliament and of the Council as regards interoperability of spatial data services, *Official Journal of the European Union*, pp. 13–34, https://doi.org/doi:10.3000/17252555.L_2010.323.eng, <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2010:323:0011:0102:EN:PDF>, 2014.
- 550 Evans, S. G. and Clague, J. J.: Recent climatic change and catastrophic geomorphic processes in mountain environments, *Geomorphology*, 10, 107–128, [https://doi.org/10.1016/0169-555X\(94\)90011-6](https://doi.org/10.1016/0169-555X(94)90011-6), <https://linkinghub.elsevier.com/retrieve/pii/0169555X94900116>, 1994.
Facebook: React, <https://reactjs.org/>, 2013.
- Friele, P., Millard, T. H., Mitchell, A., Allstadt, K. E., Menounos, B., Geertsema, M., and Clague, J. J.: Observations on the May 2019 Joffre Peak landslides, British Columbia, *Landslides*, <https://doi.org/10.1007/s10346-019-01332-2>, 2020.
- 555 Friele, P. A.: Volcanic Landslide Risk Management, Lillooet River Valley, BC: Start of north and south FSRs to Meager Confluence, Meager Creek and Upper Lillooet River, Tech. rep., 2012.
- Froude, M. J. and Petley, D. N.: Global fatal landslide occurrence from 2004 to 2016, *Natural Hazards and Earth System Sciences*, 18, 2161–2181, <https://doi.org/10.5194/nhess-18-2161-2018>, <https://www.nat-hazards-earth-syst-sci.net/18/2161/2018/>, 2018.
- Geertsema, M.: Quick Clay, pp. 803–804, Springer Netherlands, Dordrecht, https://doi.org/10.1007/978-1-4020-4399-4_282, <https://doi.org/10.1007/978-1-4020-4399-4{ }282>, 2013.
- 560 Gilpin, L. H., Bau, D., Yuan, B. Z., Bajwa, A., Specter, M., and Kagal, L.: Explaining explanations: An overview of interpretability of machine learning, *Proceedings - 2018 IEEE 5th International Conference on Data Science and Advanced Analytics, DSAA 2018*, pp. 80–89, <https://doi.org/10.1109/DSAA.2018.00018>, 2019.
- Goodfellow, I., Bengio, Y., and Aaron, C.: *Deep Learning*, MIT press., 2016.
- 565 Goudie, A.: Alphabetical Glossary of Geomorphology, *International Association of Geomorphologists*, p. 84, 2014.
- Gruber, S.: Derivation and analysis of a high-resolution estimate of global permafrost zonation, *Cryosphere*, 6, 221–233, <https://doi.org/10.5194/tc-6-221-2012>, 2012.
- Guzzetti, F., Carrara, A., Cardinali, M., and Reichenbach, P.: Landslide hazard evaluation: a review of current techniques and their application in a multi-scale study, Central Italy, *Geomorphology*, 13, 1995, [https://doi.org/10.1016/S0169-555X\(99\)00078-1](https://doi.org/10.1016/S0169-555X(99)00078-1), 1999.
- 570 Guzzetti, F., Mondini, A. C., Cardinali, M., Fiorucci, F., Santangelo, M., and Chang, K. T.: Landslide inventory maps: New tools for an old problem, *Earth-Science Reviews*, 112, 42–66, <https://doi.org/10.1016/j.earscirev.2012.02.001>, <http://dx.doi.org/10.1016/j.earscirev.2012.02.001>, 2012.
- Guzzetti, F., Gariano, S. L., Peruccacci, S., Brunetti, M. T., Marchesini, I., Rossi, M., and Melillo, M.: Geographical landslide early warning systems, *Earth-Science Reviews*, 200, 102973, <https://doi.org/10.1016/j.earscirev.2019.102973>, <https://doi.org/10.1016/j.earscirev.2019.102973>, 2020.
- 575 Hajimoradlou, A., Roberti, G., and Poole, D.: A Probabilistic Approach for Predicting Landslides by Learning a Self-Aligned Deep Convolutional Model, <http://arxiv.org/abs/1911.04651>, 2019.
- Hajimoradlou, A., Roberti, G., and Poole, D.: Predicting Landslides Using Locally Aligned Convolutional Neural Networks, in: *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, {IJCAI-20}*, edited by Bessiere, C., International Joint Conferences on Artificial Intelligence Organization, 2020.
- 580 Howes, D. E. and Kenk, E.: *Terrain Classification System for British Columbia*, Tech. Rep. Version 2, Fisheries Branch Ministry of Environment and Surveys and Resource Mapping Branch Ministry of Crown Lands Province of British Columbia, 1997.

- Hungr, O., Leroueil, S., and Picarelli, L.: The Varnes classification of landslide types, an update, *Landslides*, 11, 167–194, <https://doi.org/10.1007/s10346-013-0436-y>, 2014.
- 585 Hykes, S.: Docker, <https://www.docker.com/>, 2013.
- INSPIRE Thematic Working Group Natural Risk Zones: D2.8.III.12 Data Specification on Natural Risk Zones – Technical Guidelines, Tech. Rep. March 2007, European Commission Joint Research Centre, 2013.
- ISA: Re3gistry software 1.3, [https://ec.europa.eu/isa2/solutions/re3gistry{_\)en](https://ec.europa.eu/isa2/solutions/re3gistry{_)en), 2016.
- Jackson Jr, L., Smyth, C., and Poole, D.: Hazardmatch: an application of artificial intelligence to landslide susceptibility mapping, Howe
590 Sound Area, British Columbia, 4th Canadian Conference on Geohazards: From Causes to Management, p. 594, 2008.
- Jackson Jr, L. E.: Recommendation for adding logging, logging road, wildfire, and morphometric parameters to the soil-slide model, Tech. rep., [https://italy.minervageo.com/wp-content/uploads/2020/07/Minerva{_\)Jackson2019.pdf](https://italy.minervageo.com/wp-content/uploads/2020/07/Minerva{_)Jackson2019.pdf), 2019.
- Jakob, M.: A size classification for debris flows, *Engineering Geology*, 79, 151–161, <https://doi.org/10.1016/j.enggeo.2005.01.006>, 2005.
- Marcus, G.: Deep Learning: A Critical Appraisal, pp. 1–27, <http://arxiv.org/abs/1801.00631>, 2018.
- 595 Marcus, G. and Davis, E.: *Rebooting AI: Building Artificial Intelligence We Can Trust*, Knopf Doubleday Publishing Group, 2019.
- McDougall, S.: 2014 Canadian Geotechnical Colloquium: Landslide runout analysis — current practice and challenges, *Canadian Geotechnical Journal*, 54, 605–620, <https://doi.org/10.1139/cgj-2016-0104>, 2016.
- Mergili, M., Marchesini, I., Alvioli, M., Metz, M., Schneider-Muntau, B., Rossi, M., and Guzzetti, F.: A strategy for GIS-based 3-D slope stability modelling over large areas, *Geoscientific Model Development*, 7, 2969–2982, <https://doi.org/10.5194/gmd-7-2969-2014>, 2014.
- 600 Mergili, M., Fischer, J. T., Krenn, J., and Pudasaini, S. P.: R.avaflow v1, an advanced open-source computational framework for the propagation and interaction of two-phase mass flows, *Geoscientific Model Development*, 10, 553–569, <https://doi.org/10.5194/gmd-10-553-2017>, 2017.
- MetaCarta: OpensLayer, <https://openlayers.org/>, 2005.
- Mijić, N. and Bartha, G.: Infrastructure for Spatial Information in European Community (INSPIRE) Through the Time from 2007. Until
605 2017, pp. 34–42, https://doi.org/10.1007/978-3-030-02577-9_5, 2018.
- Minerva Intelligence: Minerva INSPIRE registry, <http://minerva.codes/registry>, 2019a.
- Minerva Intelligence: Natural Hazard Category Landslide Extension, <http://minerva.codes/codelist/NaturalHazardCategoryLandslideExtension>, 2019b.
- Minerva Intelligence: Landslide Size Class, <http://minerva.codes/codelist/LandslideSizeClass>, 2019c.
- 610 Minerva Intelligence: Aristotelean Class Editor - ACE, <https://minervaintelligence.com/project/ace/>, 2019d.
- Minerva Intelligence: INSPIRE Natural Risk Zone Schema Extension for Susceptibility Area, <https://github.com/minervaintelligence/INSPIRE-NZ-Susceptibility>, 2019e.
- Minerva Intelligence: Sea to Sky hazards, <https://www.minervageohazards.com/>, 2019f.
- Napolitano, E., Marchesini, I., Salvati, P., Donnini, M., Bianchi, C., and Guzzetti, F.: LAND-deFeND – An innovative
615 database structure for landslides and floods and their consequences, *Journal of Environmental Management*, 207, 203–218, <https://doi.org/10.1016/j.jenvman.2017.11.022>, <https://doi.org/10.1016/j.jenvman.2017.11.022>, 2018.
- Newman, J. P., Maier, H. R., Riddell, G. A., Zecchin, A. C., Daniell, J. E., Schaefer, A. M., van Delden, H., Khazai, B., O’Flaherty, M. J., and Newland, C. P.: Review of literature on decision support systems for natural hazard risk reduction: Current status and future research directions, *Environmental Modelling and Software*, 96, 378–409, <https://doi.org/10.1016/j.envsoft.2017.06.042>, [http://dx.doi.org/10.1016/](http://dx.doi.org/10.1016/j.envsoft.2017.06.042)
620 [j.envsoft.2017.06.042](http://dx.doi.org/10.1016/j.envsoft.2017.06.042), 2017.

- Pearl, J.: Probabilistic reasoning in intelligent systems: networks of plausible inference, Morgan Kaufmann, 1988.
- Phengsuwan, J., Shah, T., James, P., Thakker, D., Barr, S., and Ranjan, R.: Ontology-based discovery of time-series data sources for landslide early warning system, *Computing*, <https://doi.org/10.1007/s00607-019-00730-7>, <https://doi.org/10.1007/s00607-019-00730-7>, 2019.
- Poole, D. and Mackworth, A.: *Artificial Intelligence: foundations of computational agents*, Cambridge University Press, second edn., 2017.
- 625 Poole, D. and Smyth, C.: Type Uncertainty in Ontologically-Grounded Qualitative Probabilistic Matching, in: *Symbolic and Quantitative Approaches to Reasoning with Uncertainty*, edited by Godo, L., pp. 763–774, Springer Berlin Heidelberg, Berlin, Heidelberg, 2005.
- Poole, D., Smyth, C., and Sharma, R.: Ontology Design for Scientific Theories That Make Probabilistic Predictions, *IEEE Intelligent Systems*, 24, 27–36, 2009.
- Pudasaini, S. P.: A general two-phase debris flow model, *Journal of Geophysical Research: Earth Surface*, 117, 1–28, 630 <https://doi.org/10.1029/2011JF002186>, 2012.
- Refraction Research: PostGIS, <https://postgis.net/>, 2001.
- Reichenbach, P., Rossi, M., Malamud, B. D., Mihir, M., and Guzzetti, F.: A review of statistically-based landslide susceptibility models, *Earth-Science Reviews*, 180, 60–91, <https://doi.org/10.1016/j.earscirev.2018.03.001>, <https://doi.org/10.1016/j.earscirev.2018.03.001>, 2018.
- 635 SafeLand: Recommended procedures for validating landslide hazard and risk models and maps, *Living with landslide risk in Europe: Assessment, effects of global change, and risk management strategies*, p. 162, 2011.
- Segoni, S., Tofani, V., Rosi, A., Catani, F., and Casagli, N.: Combination of Rainfall Thresholds and Susceptibility Maps for Dynamic Landslide Hazard Assessment at Regional Scale, *Frontiers in Earth Science*, 6, <https://doi.org/10.3389/feart.2018.00085>, <https://www.frontiersin.org/article/10.3389/feart.2018.00085/full>, 2018.
- 640 Sermet, Y. and Demir, I.: Towards an information centric flood ontology for information management and communication, *Earth Science Informatics*, 12, 541–551, <https://doi.org/10.1007/s12145-019-00398-9>, <https://doi.org/10.1007/s12145-019-00398-9>, 2019.
- Shapiro, S. C.: *Encyclopedia of Artificial Intelligence*, John Wiley & Sons, Inc., USA, 2nd edn., 1992.
- Sharma, R., Poole, D., and Smyth, C.: A framework for ontologically-grounded probabilistic matching q, *International Journal of Approximate Reasoning*, 51, 240–262, <https://doi.org/10.1016/j.ijar.2009.05.007>, <http://dx.doi.org/10.1016/j.ijar.2009.05.007>, 2010.
- 645 Smith, B.: Ontology, in: *Blackwell Guide to the Philosophy of Computing and Information*, edited by Luciano Floridi, chap. 11, pp. 155–166, Blackwell, Oxford, 2003.
- Smyth, C. and Poole, D.: Qualitative Probabilistic Matching with Hierarchical Descriptions, *{KR}2004: Principles of Knowledge Representation and Reasoning*, pp. 479–486, 2004.
- Smyth, C., Poole, D., and Sharma, R.: *Semantic e-Science and Geology*, Association for the Advancement of Artificial Intelligence, 2007.
- 650 Strahler, A. N.: Quantitative Analysis of Watershed Geomorphology, *Transactions of the American Geophysical Union.*, *Transactions, American Geophysical Union*, 38, 913–920, 1957.
- The Open Planning Project: Geoserver, <http://geoserver.org/>, 2001.
- Tomas, R., Harrison, M., Barredo, J. I., Thomas, F., Llorente Isidro, M., Pfeiffer, M., and Čerba, O.: Towards a cross-domain interoperable framework for natural hazards and disaster risk reduction information, *Natural Hazards*, 78, 1545–1563, <https://doi.org/10.1007/s11069-015-1786-7>, 2015.
- 655 USDA: Chapter 3 Engineering Classification of Earth Materials, Part 631 National Engineering Handbook, p. 35, 2012.

- Van Den Eeckhaut, M. and Hervás, J.: State of the art of national landslide databases in Europe and their potential for assessing landslide susceptibility, hazard and risk, *Geomorphology*, 139-140, 545–558, <https://doi.org/10.1016/j.geomorph.2011.12.006>, <http://dx.doi.org/10.1016/j.geomorph.2011.12.006>, 2012.
- 660 Van Den Eeckhaut, M., Hervas, J., and Montanarella, L.: Landslide Databases in Europe: Analysis and Recommendations for Interoperability and Harmonisation, in: *Landslide Science and Practice: Landslide Inventory and Susceptibility and Hazard Zoning*, vol. 1, pp. 243–247, <https://doi.org/10.1007/978-3-642-31325-7>, <http://link.springer.com/10.1007/978-3-642-31325-7>{%}0A<http://www.springer.com/earth+sciences+and+geography/natural+hazards/book/978-3-642-31324-0>{%}0A<http://grass-italia.como.polimi.it/meeting2012/09marchesini.pdf>, 2013.
- 665 W3C Working Group: RDF 1.1 Primer, <https://www.w3.org/TR/rdf11-primer>, 2014.
- WeTransform: Hale Studio, <https://www.wetransform.to/products/halestudio/>, 2008.
- WeTransform: Hale Connect, <https://www.wetransform.to/products/haleconnect/>, 2014.
- Wilford, D. J., Sakals, M. E., Innes, J. L., Sidle, R. C., and Bergerud, W. A.: Recognition of debris flow, debris flood and flood hazard through watershed morphometrics, *Landslides*, 1, 61–66, <https://doi.org/10.1007/s10346-003-0002-0>, 2004.

List of all relevant changes made in the manuscript

- Added method figure and re-arranged the methods accordingly, as per suggestion of Reviewer 2
- Modified UML figure showing only the extension implemented in this project
- Deleted former Figure 5 and replaced it with Tables 4 and 5.
- Edited introduction to clarify the AI approach
- Edited the methods to clarify models, and scoring system
- Edited results and discussions
- Clarified in the discussion the main contribution of this work, which are:
 - the susceptibility schema extension
 - code list extensions
 - an example application making use of using ontologies embedded with standardized terminology, leveraging the INSPIRE spatial data infrastructure
- Added links to the model and tables showing all of the models in the appendices.
- General text clean-up

Point-by-point response to the Editor

Comments to the Author:

dear authors,

after checking the referees' reports, the discussion section and your submitted manuscript, we believe that your paper has potential interest for the journal readers but it is still in a form which is not acceptable by the journal, due to some issues that are quite important.

For this reason, we will reconsider a new, revised submission should you be ready to incorporate some major changes, as requested by the two referees.

Dear Editor, thank you for considering the manuscript for submission to the journal, we have addressed all the points raised by the reviewers and edited the manuscript accordingly.

In particular, I feel that some of their requests are particularly reasonable and, therefore, mandatory to bring the manuscript up to the journal standards. I refer to:

1. the need to clarify the methodology with respect to the (many) existing standards existing in landslide hazard studies

We re-organized the methodology and results, and a summary of landslide susceptibility mapping approaches is provided in section 1.4. A full review of landslide hazard methods is beyond the scope of this project, and as noted, the community is far from agreeing on unique standard.

Section 1.4 reads: "Landslide susceptibility is defined as the relative spatial probability of occurrence for a landslide based on the intrinsic properties of a site (SafeLand, 2011). The concept of susceptibility differs from hazard in that the temporal probability of occurrence, the triggering factors, and the magnitude of the event are not considered in the definition of a susceptibility map (SafeLand,2011; Van Den Eeckhaut and Hervás, 2012). To produce landslide susceptibility maps, three approaches are usually applied: statistical, physical, and expert-based (SafeLand, 2011). Statistical methods rely on the analysis of landslide databases and their relation to landscape properties (see review by Reichenbach et al., 2018); physical methods calculate the limit equilibrium between failure resisting forces and driving forces in slopes (e.g., Baum et al., 2008); and expert-based methods rely on expert opinion and

the assumption that influencing factors are known and are specified in the models (Dai et al., 2002). The AI approach used in this study is an example of the expert-based approach, as the models follow rules that represent the reasoning process of a landslide-expert, providing semi-quantitative susceptibility map”

2. the need to clarify, homogenize and standardize the taxonomy, the ontology and the terminology used (including the way you refer to the landslide object inside a classification method which is rather different from the ones widely accepted

I do not understand this comment. The landslide taxonomy refers to the updated Varnes classification, the ontology uses CGI UIGS rock taxonomy, British Columbia terrain mapping standards, Corinne Land cover and Strahler stream order and other standardized classifications. The susceptibility schema extension is drafted from the Safeland 2011 standards.

3. the need to detail some parts of the methodology by adding figures and documentation (e.g. flow chart and expert rules used)

We addressed this point by re-organizing the manuscript as suggested by Reviewer 2.

4. the need to try to quantify the impact of the different possible choices of rules and settings on the final result

While we certainly agree that our chosen "rules" and "settings" should be scrutinized, we emphasize that the final result of the paper is not the assessments of landslide susceptibility. Instead, the final result is the susceptibility extension in INSPIRE, as well as the development of code lists and a framework within which landslide susceptibility data can be encoded. The ontology-based landslide susceptibility assessment is intended to be used as an example of the benefits of applying the developed INSPIRE landslide framework, rather than an exhaustive assessment of landslide susceptibility. For this reason, we believe that a sensitivity analysis would be beyond the scope of this paper.

5. the authors should be more careful in stating quite uncertain and very challenging objectives, such as those of unifying EU INSPIRE standards in AI intelligence applications and terminology since those standards are still debated and non-existing at the moment. Moreover, they cannot be defined by force since they require a general agreement among the many institutions and consortiums working on landslides and related risks. The task is overwhelmingly complex and the authors do not seem to clarify what is the contribution of their work towards this general objective

Our contributions are discussed in section 4 , and we have clarified some of the language used in this section to more clearly articulate our specific contribution, which includes the following three accomplishments:

- the susceptibility schema extension
- code list extensions
- an example of the benefits of applying INSPIRE standards in the form of ontologically-driven landslide susceptibility assessments

See our reply to Reviewer 1:

In this study we start from showing how to modify INSPIRE to make it possible to use it for landslide-specific applications. By suggesting a landslide-specific schema and code list extensions, we set the ground for INSPIRE-compliant landslide susceptibility studies. Other organizations can build on top of these extensions and future landslide susceptibility applications can be compared as they formally refer to the same data structure and semantics. Note that we do not force any specific data and modeling variable selection, nor modeling approach for landslide susceptibility/hazard/risk method. Such an effort is beyond the scope to this paper and, to some extent, already addressed by the SafeLand project, rather, we provide the data structure and semantics to store and share whichever method has been chosen by the modeler. For example, data selection for calculation of landslide

susceptibility is encompassed in the schema structure under “Influencing Factor” which are “unbounded in multiplicity and can be defined qualitatively or quantitatively”, leaving a broad range of possibilities to the modeler.

6. the need to better frame the entire method and results within the existing landslide risk mitigation framework (by also citing the most relevant literature). When the proposed tools would be actually usable and useful? Possibly, where no accurate landslide inventory maps are already available? And under which conditions and pre-requisites?

As noted in the response to (1), we re-organized the methodology and results, and a summary of landslide hazard approaches is provided in section 1.4. We also clarified some of the language that specifies when the ontological approach can (or should) be used, in section 4.3.

Section 1.4: “Landslide susceptibility is defined as the relative spatial probability of occurrence for a landslide based on the intrinsic properties of a site (SafeLand, 2011). The concept of susceptibility differs from hazard in that the temporal probability of occurrence, the triggering factors, and the magnitude of the event are not considered in the definition of a susceptibility map (SafeLand,2011; Van Den Eeckhaut and Hervás, 2012). To produce landslide susceptibility maps, three approaches are usually applied: statistical, physical, and expert-based (SafeLand, 2011). Statistical methods rely on the analysis of landslide databases and their relation to landscape properties (see review by Reichenbach et al., 2018); physical methods calculate the limit equilibrium between failure resisting forces and driving forces in slopes (e.g., Baum et al., 2008); and expert-based methods rely on expert opinion and the assumption that influencing factors are known and are specified in the models (Dai et al., 2002). The AI approach used in this study is an example of the expert-based approach, as the models follow rules that represent the reasoning process of a landslide-expert, providing semi-quantitative susceptibility map”

Added in section 4.3: “The main goal of this paper is not to present the semantic matching approach, but to show an example on how to modify INSPIRE to make it possible to use it for landslide-specific applications. By suggesting these landslide-specific schema and code list extensions, we set the ground for INSPIRE-compliant landslide susceptibility studies. Other organizations can build on top of these extensions and future landslide susceptibility applications can be compared as they formally refer to the same data structure and semantics. Note that we do not force any specific data and modeling variable selection, nor modeling approach for landslide susceptibility/hazard/risk method. Such an effort is beyond the scope to this paper and, to some extent, already addressed by the SafeLand project (e.g., SafeLand, 2011) rather, we provide the data structure and semantics to store and share whichever method has been chosen by the modeler. For example, data selection for calculation of landslide susceptibility is encompassed in the schema structure under “Influencing Factor” which are “unbounded in multiplicity and can be defined qualitatively or quantitatively”, leaving broad range of possibilities to the modeler. Regarding the data quality, it is discussed in the Natural Risk Zone schema and they refer to ISO standards (INSPIRE Thematic Working Group Natural Risk Zones, 2013). However, we recognize that specific code list (semantics) dealing with data quality and model uncertainty are missing. We hope that the INSPIRE thematic group will address this point.

Please also answer to all the remaining minor issues and specific comments highlighted in the reviewer's reports, one by one, by modifying the manuscript where required and directly replying to all remarks and comments on a separate document to attach to the new modified manuscript.

The replies to all the reviewers' comments are below, in this document.

I am confident that, after that, the second review round might be expeditious and more successful.

Thank you.

Point-by-point response Reviewer 1.

In the presented work an application that tries to demonstrate the value of INSPIRE compliance in enhancing the knowledge interoperability in field of Landslide susceptibility analysis. The manuscript introduced, highlighted and discussed valuable information and critical points and current issues in mapping natural hazards using spatial data and AI. However, the reviewer still need to discuss here some points that still need to be elaborated within the text for clear the confusion with readers, especially those who are on various level of experience or scientific backgrounds.

GR: Dear reviewer, Thank you for the insightful revision and useful comments. Our replies are below your comments and we edited the text to address your observations when needed.

General comments:

Abstract:

1- What kind of enhancement the authors refer to rather than standardize the knowledge interoperability using the standard vocabularies, please clarify this.

GR: We enhance knowledge transfer and interoperability of data and data analytics as they are based on the same data structure and semantic standards.

2-"The use of INSPIRE-standardized vocabularies in ontologies that express scientific _ models promotes the adoption of the standards across the European Union and beyond" This sentence need enormous work to be carried to fulfill its objective, in your current work, how much or how far did you consider your present work contributes to this end? Taking in mind the various methods of the Risk assessment researcher: Data selection and representation, variable selection and optimization, modeling type physical to datamining...etc.

GR: In this study we start from showing how to modify INSPIRE to make it possible to use it for landslide-specific applications. By suggesting these landslide-specific schema and code list extensions, we set the ground for INSPIRE-compliant landslide susceptibility studies. Other organizations can build on top of these extensions and future landslide susceptibility application can be compared as they formally refer to the same data structure and semantics. Note that we do not force any specific "Data selection and representation, variable selection and optimization, modeling type physical to datamining...etc" for landslide susceptibility/hazard/risk method, rather, we provide the data structure and semantics to store and share whichever method has been chosen by the modeler. For example, data selection is encompassed in the schema structure under "Influencing Factor" which are "unbounded in multiplicity and can be defined qualitatively or quantitatively", leaving broad range of possibilities to the modeler.

To clarify this point, we added this paragraph in section 4.3: "The main goal of this paper is not to present the semantic matching approach, but to show an example on how to modify INSPIRE to make it possible to use it for landslide-specific applications. By suggesting these landslide-specific schema and code list extensions, we set the ground for INSPIRE-compliant landslide susceptibility studies. Other organizations can build on top of these extensions and future landslide susceptibility applications can be compared as they formally refer to the same data structure and semantics. Note that we do not force any specific data and modeling variable selection, nor modeling approach for landslide susceptibility/hazard/risk method. Such an effort is beyond the scope to this paper and, to some extent, already addressed by the SafeLand project (e.g., SafeLand, 2011) rather, we provide

the data structure and semantics to store and share whichever method has been chosen by the modeler. For example, data selection for calculation of landslide susceptibility is encompassed in the schema structure under "Influencing Factor" which are "unbounded in multiplicity and can be defined qualitatively or quantitatively", leaving broad range of possibilities to the modeler. Regarding the data quality, it is discussed in the Natural Risk Zone schema and they refer to ISO standards (INSPIRE Thematic Working Group Natural Risk Zones, 2013). However, we recognize that specific code list (semantics) dealing with data quality and model uncertainty are missing. We hope that the INSPIRE thematic group will address this point.

3- . As matter of fact, any analytical model using spatial data, suffers from uncertainty started with modeling ambiguity, surrogate data, error propagation..etc. in different extend, Therefore, the extension to Natural Risk zone susceptibility schema, need to consider the model accuracy assessment, validation and error quantification of data (spatial and aspatial) and used model.

GR: Data quality standards are discussed in the Natural Risk Zone schema and they refer to ISO standards (Section 7 and 8 in D2.8.III.12 Data Specification on Natural Risk Zones). However, we recognize that specific code list (semantics) dealing with data quality and model uncertainty are missing. We hope that the INSPIRE thematic group will address this point. We briefly mention the importance of models parameters semantics at lines 329-331 "by embedding the ontology concepts related to statistical parameters (e.g. receiving operating curves, confidence intervals) or physical parameters (e. g. friction angles, viscosity), it will be possible for the numerical outputs of quantitative methods to be explained in natural language"

1 introduction

L27 P2 : "EU countries are aligning and Europe (Cho and Cromptvoets, 2019)." Most of the high quality sensors collected data and field works supported by scientists located in Europe, thus, Please mention the main rebuttals why the application still limited in literature.

GR: In-depth discussion on why INSPIRE is slowly adopted is beyond the scope of this paper. Cho and Cromptvoets, (2019) suggest that the slow INPIRE adoption by EU countries may be due to legal and policy issues. Regarding the scientific literature, there are a few cases which make use of INSPIRE, and we discuss them in the paper. Furthermore, INSPIRE is a geospatial framework which is not something commonly discussed in the geological/geomorphological literature.

L43 P2: More standards are still missing need to be mentioned here, like: 1- Standards for input data volume concerning the study area extent and landslide type. 2- Standard for outcomes accuracy concerning the type/quality/amount and extent of reserch details coverage.

GR: This is a general introduction to standards and AI, beyond the field of natural hazards. Regarding the "1- Standards for input data volume concerning the study area extent and landslide type and 2- Standard for outcomes accuracy concerning the type/quality/amount and extent of research details coverage". These are a currently discussed topics in the geological/geomorphological literature, and there is no universal agreement on such standards. This paper focuses on the INSPIRE semantic and data framework standards in which multiple different approaches to landslide susceptibility mapping can fit and can provide interoperable results.

L45 P2 : "Deep learning techniques....such as JPEG, or WAV)" Deep learning still not widely applied in Hazard and vulnerability mapping. The authors may mentioned the most common models in the literature instead like physical or datamining statistical models.

GR: This is a general introduction to standards and AI, beyond the field of natural hazards. The methods to assess landslide susceptibility are discussed in section 1.4 of the paper, but a detailed review of physical and statistical methods for landslide susceptibility mapping is beyond the scope of this paper.

Section 1.4: "Landslide susceptibility is defined as the relative spatial probability of occurrence for a landslide based on the intrinsic properties of a site (SafeLand, 2011). The concept of susceptibility differs from hazard in that the temporal probability of occurrence, the triggering factors, and the magnitude of the event are not considered in the definition of a susceptibility map (SafeLand,2011; Van Den Eeckhaut and Hervás, 2012). To produce landslide susceptibility maps, three approaches are usually applied: statistical, physical, and expert-based (SafeLand, 2011). Statistical methods rely on the analysis of landslide databases and their relation to landscape properties (see review by Reichenbach et al., 2018); physical methods calculate the limit equilibrium between failure resisting forces and driving forces in slopes (e.g., Baum et al., 2008); and expert-based methods rely on expert opinion and the assumption that influencing factors are known and are specified in the models (Dai et al., 2002). The AI approach used in this study is an example of the expert-based approach, as the models follow rules that represent the reasoning process of a landslide-expert, providing semi-quantitative susceptibility map"

L210 to 221 P 10: too general information, please discuss in more details the susceptibility modeling like in light of using machine learning (classification/clustering..etc) or physical model of slope structure analysis.

GR: In this paragraph we present a detailed description of the schema structure and how inputs and outputs of susceptibility modelling can be mapped to this schema. We do not discuss how susceptibility modelling can be done, leaving space to the many possible approaches. A overview of physical and statistical methods for landslide susceptibility mapping is discussed in section 1.4 of the paper.

Section 1.4: "Landslide susceptibility is defined as the relative spatial probability of occurrence for a landslide based on the intrinsic properties of a site (SafeLand, 2011). The concept of susceptibility differs from hazard in that the temporal probability of occurrence, the triggering factors, and the magnitude of the event are not considered in the definition of a susceptibility map (SafeLand,2011; Van Den Eeckhaut and Hervás, 2012). To produce landslide susceptibility maps, three approaches are usually applied: statistical, physical, and expert-based (SafeLand, 2011). Statistical methods rely on the analysis of landslide databases and their relation to landscape properties (see review by Reichenbach et al., 2018); physical methods calculate the limit equilibrium between failure resisting forces and driving forces in slopes (e.g., Baum et al., 2008); and expert-based methods rely on expert opinion and the assumption that influencing factors are known and are specified in the models (Dai et al., 2002). The AI approach used in this study is an example of the expert-based approach, as the models follow rules that represent the reasoning process of a landslide-expert, providing semi-quantitative susceptibility map"

L275 P 17: "Quantification of this value has yet to be made, but calculations.. Lombardia Region, Italy, alone." Please extend this paragraph by mentioning some details or statistics.

GR: We realized that sentence was incorrect and rephrased: "A comparative study (Craglia and Campagna, 2010) of regional SDI in the context of INSPIRE implementation, showed that inefficient data access and use at the European level results in economic losses in the 100-200 Million Euro annual range. The same study, shows that the regional SDI of Lombardia, Italy, allowed € 3 m/year savings to companies working in Environmental Impact Assessments (EIA), and Strategic Environmental Assessments (SEA). Savings in the same order of magnitude can be expected by adopting INSPIRE standards in the geological hazard assessment domain."

L285 to L315 P 18: Can we assume that, the mentioned limitations, were the reason behind the scare mentioned in L28,29 P2 " EU countries are aligning and serving

INSPIRE data at a slow pace...are available across Europe"?

GR: The mentioned limitations are part of the reasons. Other reasons maybe that is a lot of work to align the data and the return of money and time investments for data transformation is not immediately quantifiable

L323, 324 L19: " However, in current geological assessments, expert judgment is still widely applied (e.g., Association of Professional Engineers and Geoscientists of British Columbia, 2010)," Please add recent references to support this statement.

GR: Unfortunately, guidelines for professional practice are not updated very often. In this context a reference from 2010 is to be considered "recent". We edited the text and added another reference. Now reads: "However, in current professional geological assessments and geomorphological mapping applications, expert judgment is still widely applied (e.g., Association of Professional Engineers and Geoscientists of British Columbia, 2010, Guzzetti et al., 2012),"

Specific comments: L55 to L60 P3: Not clear, please re-write.

GR: We rephrased: "For example, "Slides in soil" and "Slides in rock" share the same parent concept "Slides" and they are differentiated by the property dealing with the material type, "Soil" and "Rock", which make them uniquely identifiable. Taxonomies based on Aristotelian definitions support multi-hierarchical knowledge networks and can be used by computers to make logical inferences (Poole et al., 2009; Smith, 2003). The term 'multi-hierarchical' implies that there is more than one way to move through a taxonomy to arrive at a particular node or term. For example, the landslide taxonomy can be arranged based on different properties. If the landslide types are firstly arranged based on the type of movement and then based on the type of material, one path within the taxonomy would be: Landslide> slides> slides in rock and slides in soil. Alternatively, if the landslide types are arranged first based on the material type and then on the movement type, the path of the taxonomy would be: Landslide> landslides in rock> slides in rock and flows in rock. Both paths are valid, but they reach the same concept in different ways

L125 P3: "in geohazard mapping to produce landslide susceptibility maps (Jackson Jr et al., 2008)" Can you add more recent studies.

GR: There is no other recent study using this expert-based approach based on ontological matching for landslide susceptibility mapping. The framework presented in this paper can be adopted with any method used to assess landslide susceptibility. In the schema, landslide susceptibility is an element that can be quantitatively or qualitatively defined.

Figure 2. Please mention one study for each sub-category

GR: I do not understand this comment. Do you mean a reference for each landslide type? They are mainly from Hungr et al 2014.. details on the properties used for the classification are in Appendix B - Properties used for the landslide classification

Figure 3 and 5. Please use multiple layout as the figure is not readable on A4 paper mode.

GR: We modified figure 3, showing only the extension done for this project, and we deleted Figure 5 and replace it with Tables 4 and 5. In the tables captions there is the link to the actual match report table from the webmap application

Point-by-point response to Reviewer 2.

1 Introduction

Reading this manuscript was interesting and stimulating. The article deals with the problem of landslide susceptibility mapping by combining different aspects, ranging from (i) the definition of a landslides classification compliant with the INSPIRE Directive, (ii) the definition of a multi-hierarchical model for the same classification, (iii) the definition of an INSPIRE scheme for landslide susceptibility and (iv) the definition of an expert-based method for the generation of maps of susceptibility to specific types of landslides in areas where inventories are scarce. In this sense the paper is adequate to the journal also because it presents innovative concepts which are adequate to international standards. Despite this I think that interventions are needed to improve the paper and make it suitable to be published.

GR: Dear reviewer, thank you for your careful revision and your insightful comments. We have addressed point by point your comments and updated the text accordingly.

2 General comments

The manuscript makes extensive use of the term artificial intelligence (also in the title). However, the type of approach used is not the classical application of artificial intelligence expected from the literature. I think it is important to clarify why this terminology is used (using the literature) or alternatively to modify the title and the text focusing more on ontology and taxonomy. In the latter case some of the titles used for the different sections of the manuscript should be changed.

GR: As stated in line 34, "Artificial intelligence" is "the synthesis and analysis of computational agents that act intelligently" (Poole and Mackworth, 2017). This definition encompasses a broad range of methods and algorithms including, but not limited to, machine learning. I understand that commonly the term artificial intelligence is used as synonym for machine learning, which in turn is a term used to talk about various statistical methods. We rephrased the paragraph.

We rephrased the paragraph to clarify this point: "The term Artificial Intelligence is commonly used to indicate only the machine learning part of the field, especially in the landslide literature (e.g., Dieu and Gjermundsen, 2020). In this paper we use the term in its broader connotation, which includes also the ontological method used in this paper. See below for the description of the method and definition of ontologies"

In the Method section, I think that there is a need of an introductory sub-sections which briefly introduce the flowchart of the method, even using a figure. I suggest something like: (i) defining taxonomies, (ii) defining expert-based rules, (iii) performing matching, (iv) deriving the susceptibility map. As a consequence of the flow-chart introduction the subsequent sub-sections could use a title which is compliant with the flowchart content.

GR: Good point, thank you. We have added a flowchart figure and updated the methods adding an intro. We also have rearranged Methods and Results to remove some of the repetition and add details in the webmap final results.

Sub-section 3.1.1 describe the creation of the NH classification code list for landslides. It was used for defining the expert-based rules. I wonder if this sub-section should be moved in the method section. Moreover it could be helpful to describe how the flat representation of your classification differs from the classification of Hungr et al. (2014)

GR: We reword methods and results to clarify this point: "The Natural Hazard Classification code list extension for landslides considers material type and failure movement, splitting the tree, first on type of movement, and then on type of material, following Hungr et al. (2014) (Figure 3). Other properties, such as: water content, depth of failure, rate of movement, loading state, channelized state, and failure plane geometry (see Appendix B) are used to describe the individual landslide types, as the unique combination of these properties allows for unambiguous classification in an Aristotelian taxonomy. We used these properties because, even if not shown in the final taxonomic tree, they are explicitly applied in the wordy description of landslides type by Hungr et al. (2014).

The method used for building the susceptibility map is based on the definition of the rules and setting of the matching scores. I wonder if these scores can have a huge impact on the final calculation of the map. I think that a discussion about how the scores are defined and which is the effect of changing those values is needed. Looking at the matching score table it seems that unmatched means -10. Is this something that should be added in the text?

GR: You are right, these scores determine the final calculation of the map. While we certainly agree that our chosen "rules", "settings" and "scores" should be scrutinized, we emphasize that the final result of the paper is not the assessments of landslide susceptibility. Instead, the final result is the susceptibility extension in INSPIRE, as well as the development of code lists and a framework within which landslide susceptibility data can be encoded. The ontology-based landslide susceptibility assessment is intended to be used as an example of the benefits of applying the developed INSPIRE landslide framework, rather than an exhaustive assessment of landslide susceptibility. For this reason, we believe that a sensitivity analysis would be beyond the scope of this paper

That said, a detailed description of the scores is provided in lines 115-125; they are a measure of surprise that uses order of magnitude numbers to distinguish qualitative measures. To better explain and address the -10, we added in the text at line 125: "In this study, an Exact match or an AKO exact match of a property with frequency "always" scores 10000, "usually" scores 9000, "sometimes" scores 1000, "rarely" scores "100" and "never" scores -10000; unmatched attributes are awarded -10 points. These scores are an arbitrary representation of degree of surprise that uses order of magnitude numbers to distinguish qualitative measures.

For more extensive review we refer Smyth and Poole 2004:

- *always a proposition is "always" true, means that you are very surprised if it is false. All experience leads you to believe that it is always true, but you are leaving open the possibility that it isn't true*
- *usually a proposition is "usually" true means you are some- what surprised if it is false.*
- *sometimes a proposition that is "sometimes" true, means you wouldn't be surprised if it is true or false.*
- *rarely a proposition is "rarely" true means you are some- what surprised if it is true.*
- *never a proposition is "never" true, means that you are very surprised if it is true.*

In terms of the kappa calculus, we choose numbers $\alpha > 0$ and $\beta > 0$ so that:

- *always p means $\kappa(\neg p) = \alpha$ and $\kappa(p) = 0$. Thus α is the measure of surprise that p is false.*
- *usually p means $\kappa(\neg p) = \beta$ and $\kappa(p) = 0$. Thus β is the measure of surprise that p is false. The relative surprises means that $\beta < \alpha$.*
- *sometimes p means $\kappa(\neg p) = 0$ and $\kappa(p) = 0$. We are not surprised if p is true or not.*
- *rarely p means $\kappa(\neg p) = 0$ and $\kappa(p) = \beta$. Thus β is the measure of surprise that p is true. Note that "rarely" is the dual of "usually."*
- *never p means $\kappa(\neg p) = 0$ and $\kappa(p) = \alpha$. Thus α is the measure of surprise that p is true. Note that "never" is the dual of "always."*

Note that these qualitative uncertainties are only the input values (i.e., as part of the models); on output we give a numerical score (both a raw score as well as a percent of the best match). This finite scale is not adequate to describe the level of matches. In our applications we have used $\beta = 1000$ as the value for being very surprised. (The only significant as the value for being somewhat surprised and $\alpha = 10000$ feature of the values is the 10-fold ratios between them; 10 "somewhat surprised" is equal to one "very surprised").

Different expert models would lead to different susceptibility maps.

The set of rules used for defining the expert-based model are important and should be visible. Maybe I'm wrong but I can't find a table or a web address where those rules can be observed. I think that for a reader it could be very helpful to have access to those set of rules.

GR Links to the models are in appendix C. We also added them in the text in section 2.2.1.

We added in section 2.2.1: "These landslide models are intended to be proof-of-concept of models that can be used in the semantic approach proposed in this paper. In particular, some of the properties used in the models are drafted from literature analysis of logging-related landslides in British Columbia, Canada"

In the conclusions you stated that in this study you present some landslide susceptibility maps. I would rather say that that you present a method for building a landslide susceptibility map using taxonomy and predictors/covariates and that this method is useful where accurate landslides inventories are not available.

GR: Yes, we agree. We rephrased the first sentences of the conclusions "This study presents an AI-based method to produce landslide susceptibility maps using an ontology and standardized taxonomies within the framework provided by the INSPIRE Natural Risk Zone theme. This method does not need an accurate landslides inventory to make predictions. We produced susceptibility maps for debris flow, slides in soil and slides in rock for the province of Veneto, Italy."

About figures and tables: figure 5 is not clear. I would transform it to a table and I would let the user to go and see it in the website putting a link in the caption.

GR: Good point. We made the figure into two tables (4 and 5), added the link in the tables' captions, and updated the text

Specific comments

Row 45: I would remove the reference to jpeg or wav.

GR: We rephrased to: "commonly specified in data storage standards such as JPEG, or WAV"

Row 50: I think that the concept of entities and relationships should be defined a priori or some literature should be cited.

GR: We expanded the sentence and added a reference: "In particular, an ontology defines the vocabulary for individuals and relationships within a knowledge domain. Individuals may be concrete entities (e.g. a rock), or abstract concepts, (e.g. numbers); relationships are properties that describe how individuals are connected. Typical examples of relationships include: is-a-kind-of, is-part-of, is-superclass-of, has-some-property; the ontology also defines axioms controlling the use of the vocabulary for logical and thematic consistency Poole2017"

Rows 52 -54: This sentence is not immediately clear. An example could help the reader.

GR: We added an example: "For example, "Slides in soil" and "Slides in rock" share the same parent concept "Slides" and they are differentiated by the property dealing with the material type, "Soil" and "Rock", which make them uniquely identifiable. Taxonomies based on Aristotelian definitions support multi-hierarchical knowledge networks and can be used by computers to make logical inferences (Poole et al., 2009; Smith, 2003)"

Row 81: In my opinion the approach is based on the domain-expert reasoning, since rules are defined a priori. Then it mimics these rules.

GR: This is correct. The method applies rules that are set to follow expert reasoning. We rephrased to better explain: "The AI approach used in this study is an example of the expert-based approach, as the models follow rules that represent the reasoning process of a landslide-expert, providing semi-quantitative susceptibility map"

At rows 104-105 the definitions of the model is too cryptic. I would add some examples to make clear to the reader that, if I understand correctly, a model is a set of rules defined a priori by the expert and based on the properties of the entities defined in the

taxonomy.

GR: You are correct, we rephrased and added links to the expert models in section 2.2.1 and tables in appendix C.

“A model is a set of rules defined a priori by an expert, based on scientific literature, making use of the entities and properties defined in the ontology. These models aim to represent expert conceptualized descriptions of a given phenomenon or entity (e.g. landslide susceptibility).”

Row 110: semantic triple format and semantic network. Please cite a reference or define them.

GR: We rephrased: “Semantic networks are a graph representation of knowledge where nodes are concepts and edges are the semantic relation between concepts (Shapiro, 1992)”

Row 111: “revered” or “referred” ?

GR: Sorry for the typo, we fixed it

Rows 135-136: I would use “channels” and “steep channels” in place of “streams”

GR: Ok, text is updated

Row 160: About the stream line vector layer. I suppose that only the segments in in the mountain valleys were used. But what about the starting point of these segments? This is relevant since debris flows can be triggered also in the upper part of the watersheds where channels may not have been delineated. You can discuss this point for completeness.

GR: Stream line vector layer is from the Veneto Region geoportal. The dataset is not great: many channels, especially in the initiation zones, are not represented. But the idea was to use as much as possible data already available, rather than making new data. We want to provide a framework for interoperable landslide susceptibility mapping: many other methods, better data etc could be used to assess landslide susceptibility, but still can be delivered using this schema and code list extension

Row 167: is r.avafLOW suitable also for slides in rock?

GR: We think so: r.avafLOW allows to model fluid and solid fraction separately, you can play around with the parameters and recreate the runout of wide range of landside types, including slides in rock, assuming that the rock mass disintegrate and starts to behave as a flow-like landslide.

Rows 224-226: please use an example to describe what you have done to align the data to INSPIRE standards and explain why the same was not done for the other datasets (lakes, watersheds, etc).

GR: We rephrased the paragraph and added a figure showing the Hale Studio user interface: “For this study, we used open access datasets from the Veneto Region Geoportal and other sources (Table 2 and 3). Aligning all input datasets was beyond the scope of this project. We did, however, want to show the value of INSPIRE-aligned data and therefore aligned stream network, CORINE land cover, bedrock geology, and the Italian Landslide Inventory (IFFI) (Table 2) to INSPIRE using Hale Studio (WeTransform, 2008). Figure 5 shows how different tools in Hale Studio are used to align properties from the source dataset to the target dataset. For example, the field “eta” – “Age” in Italian, of the original Veneto dataset, was directly mapped to four different INSPIRE fields: the olderNamedAge.href and title and the youngerNamedAge.href and title. Note that olderNamedAge.href youngerNamedAge.href are hyperlinks to the code list value id and the title is the actual code list term from the GeochronologicEraValue code list. This alignment is done with many classification methods, including: Groovy Scripts, formatted strings and assign-alignment tools. For further explanation on term alignments, refer to the documentation of Hale Studio (WeTransform, 2008). Datasets used that were not compliant with INSPIRE include: lakes, watersheds, permafrost, fire, slope angle, faults, soil, roads and railways (Table 3).

Table 2: please explain how you have used the IFFI database in your model. I suppose it was used for helping to define the expert-based rules.

GR: Yes, it is correct. Expert-based model states that: Slides in Soil - has landslide type - slides in soil - always. As landslides are more likely to occur on slopes or valleys that have experienced landslides before. See Appendix C.

Table 3: how the Watersheds, Railroad and Road layers have been used in the model?

GR: Watersheds have been used to with Melton Ratio to classify catchment as debris flow, debris flood and flood prone

Roads and rail roads have been used by assuming that roadcut and railroad cut affect slope stability, when compared to undisturbed slopes. See appendix C.

Rows 231-234: what about the slope map? Was the map expressed in terms of classes of slope and not in degrees or percentage?

GR: Yes, slope classes are based on degrees. The matching systems adopted in this paper requires discretized data, which can be a benefit, as numbers (e.g. 37°) are harder than words (e.g. steep) to understand by non-technical people

Row 245: 99,9th percentile of the slope units susceptibility values. I suppose. Please specify it.

GR: Yes, correct, we rephrased to: "99,9th percentile score (i. e. susceptibility values) of instances (slope units and stream buffer polygons) for each landslide type"

Row 275: please remove brackets

GR: That sentence was actually incorrect. We rephrased: "A comparative study (Craglia and Campagna, 2010) of regional SDI in the context of INSPIRE implementation, showed that inefficient data access and use at the European level results in economic losses in the 100-200 Million Euro annual range. The same study, shows that the regional SDI of Lombardia, Italy, allowed € 3 m/year savings to companies working in Environmental Impact Assessments (EIA), and Strategic Environmental Assessments (SEA). Savings in the same order of magnitude can be expected by adopting INSPIRE standards in the geological hazard assessment domain."