"Contextualized VGI" creation and management to cope with uncertainty and imprecision

Gloria Bordogna, Luca Frigerio, Tomas Kliment, Pietro Alessandro Brivio, Laure Hossard, Giacinto Manfron, Simone Sterlacchini

HAL Id: hal-02637589

https://hal.inrae.fr/hal-02637589

Submitted on 27 May 2020

Article

"Contextualized VGI" Creation and Management to Cope with Uncertainty and Imprecision

Gloria Bordogna [1,*]**, Luca Frigerio** [2]**, Tomáš Kliment** [2]**, Pietro Alessandro Brivio** [1]**,
Laure Hossard** [3]**, Giacinto Manfron** [1] **and Simone Sterlacchini** [2]

[1]   Consiglio Nazionale delle Ricerche—Istituto per il Rilevamento Elettromagnetico dell'Ambiente
       (CNR—IREA), 20133 Milan, Italy; brivio.pa@irea.cnr.it (P.A.B.); manfron.g@irea.cnr.it (G.M.)
[2]   Consiglio Nazionale delle Ricerche—Istituto per la Dinamica dei Processi Ambientali (CNR—IDPA),
       20126 Milan, Italy; luca.frigerio@idpa.cnr.it (L.F.); tomas.kliment@idpa.cnr.it (T.K.);
       simone.sterlacchini@idpa.cnr.it (S.S.)
[3]   Institut National de la Recherche Agronomique (INRA) UMR 951 Innovation, 2 Place Pierre Viala,
       34060 Montpellier CEDEX 2, France; laure.hossard@inra.fr
*   Correspondence: bordogna.g@irea.cnr.it; Tel.: +39-02-23699-287

Abstract: This paper investigates the causes of imprecision of the observations and uncertainty of the authors who create Volunteer Geographic Information (VGI), i.e., georeferenced contents generated by volunteers when participating in some citizen science project. Specifically, various aspects of imprecision and uncertainty of VGI are outlined and, to cope with them, a knowledge-based approach is suggested based on the creation and management of "contextualized VGI". A case study example in agriculture is reported where contextualized VGI can be created about in situ crops observations by the use of a smart app that supports volunteers by means of both an ontology and the representation of the context of the geo-localization. Furthermore, an approach to cope with both ill-defined knowledge and volunteer's uncertainty or imprecise observations is defined based on a fuzzy ontology with uncertainty level-based approximate reasoning. By representing uncertainty and imprecision of VGI, users, i.e., consumers, can exploit quality checking mechanisms to filter VGI based on their needs.

Keywords: contextualized Volunteered Geographic Information (VGI); smart app; uncertainty; Imprecision; fuzzy ontology; level-based approximate reasoning

1. Introduction

The contribution of volunteers to science is an old practice that helped enrich museum collections over the years. Nowadays, this old practice has become very popular with the name of "citizen science", empowered by the use of the Internet and both Web and smart applications that allow a huge number of volunteers to help scientists in carrying out specific tasks in scientific projects [1].

Most citizen science projects exploit Volunteer Geographic Information (VGI) [2], where volunteers are asked to provide information of various forms and nature, such as textual notes, pictures, measurements of properties relative to target objects, classes, video, and sounds, by associating a geographic reference with their observations, named geofootprint [1].

Nevertheless, many researchers are critical of the usability of VGI because it is often characterized by unknown and heterogeneous levels of uncertainty and imprecision: critical causes are the heterogeneity of both expertise and commitment of the VGI authors; the unknown conditions in which VGI was created and its intended purposes, i.e., VGI context, the heterogeneous semantics of VGI (for example due to the authors' expertise or the lack of syntactical/semantic control on the authors' entered data); and the redundancy and sparseness of VGI [3].

Three main categories of approaches are applied in citizen science projects for regulating, constraining, filtering and correcting the data entered by volunteers so as to reduce as much as possible their uncertainty, ambiguities, incompleteness and inaccuracy [3,4]:

- The ex-ante approaches have the objective of providing the volunteers with easy to access resources for a correct and constrained creation of VGI and automatic mechanism for controlling the data entry [5]. These resources include templates with automatic error checking capabilities to make the adoption of better data creation practices easier, controlled vocabularies, geographic gazetteers and ontologies in the specific scientific domain of the project.
- The ex-post approaches have the objective of cleaning and improving VGI once it is already created and stored before either publishing it on the Web or using it for scientific purposes, such as object geo-localization. A common approach is based on either cross-referencing VGI with authoritative information from administrative and commercial datasets, or annotating it by the use of gazetteers and ontologies in the specific scientific domain [6].
- Hybrid approaches use both ex-ante and ex-post techniques.

The use of ontologies is particularly diffused in citizen science projects that ask volunteers to identify objects of interest and to categorize them based on the provided domain ontology. These are the majority of the citizen science projects in astronomy, such as the *Galaxy Zoo* project that asks volunteers' to classify the shapes of hundreds of thousands of galaxies present in the huge amount of deep field images supported by the Hubble Atlas of galaxies; in emergency management projects, where citizens are asked to label with categories critical situations of which they have been testimonies [7]; and in natural sciences, such as the many projects on birds watching and plants observation. These latter projects demand high precision of the classification of the observed objects. For example, *inaturalist* [8] provides visual ontologies in the form of pictures of the species to help volunteers to create correct classifications of observations, for which they have to upload a photo that is then submitted to the manual revision of the community. Nevertheless, even with the support of the ontologies, many volunteers are doubtful about their classifications and many of them explicitly ask for community review. This uncertainty can depend on the "vagueness" of the classification criteria, the volunteer low expertise and the context of the observation.

In this paper, we propose an approach to create, represent and manage "contextualized VGI" in order to allow VGI authors specifying the context of their observations on the producers' side, and making aware the VGI consumers of the embedded uncertainty and imprecision in VGI items they analyze on the users' side.

Specifically, we will discuss the direct experience we had within the Space4Agri project whose aim was the improvement of the agronomic sector in Northern Italy [9]. Space4Agri exploits multisource information, among which remote sensing data products and VGI on crops observations created through the use of a smart app by volunteers (farmers, agronomists and citizens). The application supports the volunteer in selecting the correct crop type and its phenological stage that is related to its crop development, by a specific agronomic ontology (named Biologische Bundesanstalt, Bundessortenamt and CHemical industry scale, for short BBCH scale) [10,11]. Nevertheless, even with the help of this ontology, volunteers were faced with many uncertainties in identifying both the correct crops and their actual stages.

This experience motivated our proposal of defining a knowledge-based approach to support the ex-ante creation of contextualized VGI so as to be able to represent its context, comprising both the author self-assessment of the certainty of the observation when creating VGI and the description of the conditions of the observation.

In the paper, we first introduce the problem of uncertainty and imprecision of VGI. Subsequently, we describe the Space4Agri smart app supporting contextualized VGI creation, in its first version, by means of a classic ontology, and then by describing the lessons learned based on the users' experience that outlined the need of dealing with observation uncertainty when creating semantic annotated

VGI. Then, we discuss the inadequacy of the approaches based on classic Web Ontology Language (OWL) [12,13] and introduce the notion of fuzzy OWL-DL (Description Logic) ontology to represent ill-defined knowledge in a domain [14,15]. Finally, we propose a fuzzy ontology approach with uncertainty level-based approximate reasoning to deal with the uncertainty of volunteers when creating VGI by discussing the case study of the Space4Agri smart app.

2. Uncertainty and Imprecision of VGI

Four main causes of VGI uncertainty and imprecision can be identified:

- Incomplete or inadequate knowledge of the volunteer participating in the project: this may be particularly impacting in citizen science projects involving as volunteers the general public without providing sufficient training facilities and information on the scope of the project.
- Vagueness of the criteria provided to support volunteers in correctly describing, classifying or tagging their observations. This may happen in domains where volunteers are asked to tag the observed objects based on vague descriptions of category prototypes; examples of vague descriptions of roses and sparrows are the following: "*Roses are flowers which vary in size and shape and are usually large and showy, in colours ranging from white through yellows and reds*" and "*Sparrow are small birds with thick bills for eating seeds, and are mostly coloured grey or brown*" [8].
- Limitations of the means of observation of the volunteers: this may depend on the context of the observation such as weather conditions, distance from the object, or means of observation and measurement tools. For example, given precise descriptions of both a common nightingale as "*common nightingale is slightly larger than the European robin, at 15–16.5 cm (5.9–6.5 in) length. It is plain brown above except for the reddish tail*" and of a robin redbreast as" *robin redbreast is around 12.5–14.0 cm (5.0–5.5 in) in length, with a reddish breast and brown upperparts*", one may find it difficult to tell them apart when observed from a far position.
- Missing information on the created VGI on the consumer side, such as the lack of appropriate metadata describing the VGI creation conditions, lack of information on the imprecision of the geo-localization of VGI, etc.

From the above, it is clear that the problem of VGI uncertainty and imprecision is strictly related to the representation of both VGI semantics and context of its creation. To enable semantic interoperability of VGI applications, Bakillah et al. [16] proposed a VGI semantic data model that defines the characteristics of VGI from distinct applications, in terms of the possible types of data that might be requested to volunteers. The VGI data model is intended for ex-post semantic integration of VGI, when an application already exists, by providing the conceptual framework for the generation of common descriptions of the heterogeneous VGI data sets. These descriptions should act as common interfaces to enable the querying and correct interpretation of VGI, provided by distinct applications, through a single platform. As pointed out in [3], in order to define the VGI semantic data model, the authors of [16] reviewed some VGI citizen science initiatives and classified them according to the type of contributions, namely, VGI provided by sensor devices, geo-referenced text, or geo-referenced features, i.e., abstraction of real world phenomena such as objects (e.g., buildings and Points Of Interest (POIs)) or events. In this approach, to enable the accessibility of VGI, each VGI application must register in a service by defining an instance of the VGI data model through semantic annotation [17]. This approach may force the interpretation of the original VGI to take a specific meaning depending on the subjectivity of the single annotator, while multiple volunteers with heterogeneous characteristic may have created it with distinct scopes and interpretations.

Complementary approaches have suggested applying ex-ante semantic mechanisms to support volunteers in creating VGI meaningful for the context of an application. Ex-ante semantic support for creating VGI has been recently faced in two distinct ways: either by adopting the Semantic Web and Linked Data framework [18–20] or by relying on domain ontologies [21–23]. To this end, they provide a contextual description of the meaningful concepts and relationship used in the application to benefit

not only VGI authors but also VGI users. Forcing volunteers to select tags from a vocabulary of meaningful terms and relationships, and possibly complementing the tags with free text and pictures, can be considered as a kind of ex-ante semantic annotation directly performed by VGI authors. We believe that this approach is much more respectful of the original intended meaning of VGI than what the ex-post semantic annotation can do. Nevertheless, when creating VGI from a mobile device, not connected to the Internet, Linked Data cannot be accessed. This is the main reason that favors the adoption of ontologies to support ex-ante semantic annotation of VGI during its creation [21]. Besides, during the analysis phase, the ontology provides useful complementary information on the observations. These are the reasons that make ontologies widely used in citizen science initiatives. We mention the citizen science projects under the *inaturalist* umbrella [8], supplying a visual taxonomy to support identification of plants and animal species, and the possibility to specify the imprecision of the geofootprint where the objects are observed (see Figure 1). Nevertheless, citizens often find it difficult to correctly classify objects based on the use of classic ontologies. In [7], VGI produced by volunteers via the *Ushahidi* web platform in response to the earthquake of 2010 in Haiti were examined to assess their uncertainty. Volunteers translated messages (text, e-mail, and voice) submitted by victims of the earthquake and categorized each message by the help of a two level taxonomy into a primary and secondary category and sub-category expressing distinct "emergency need" and georeferenced each message. The analysis illustrated that 50% of the messages were mis-categorized by the volunteers, failing to convey the main idea of the victim's message. At the subcategory level, approximately 73% of the messages failed to convey the main idea of the messages. Notwithstanding these results, Camponovo and Freundschuh [7] recognized the utility of VGI for emergency rescue and outlined the need for novel methods to reduce the uncertainty in VGI creation. On the other side, Haklay [1] suggests a complementary view to deal with the uncertainty of VGI in citizen science projects. The author states that VGI uncertainty should not be regarded as something that can be eliminated by using tighter protocols for its creation, but as an integral part of any VGI collection, and thus he advocates novel methods to represent and to deal with VGI uncertainty during the analysis phase. This is the point of view from which we started to design a smart app for VGI creation within the Space4Agri citizen science project capable to represent and manage uncertainty and imprecision.
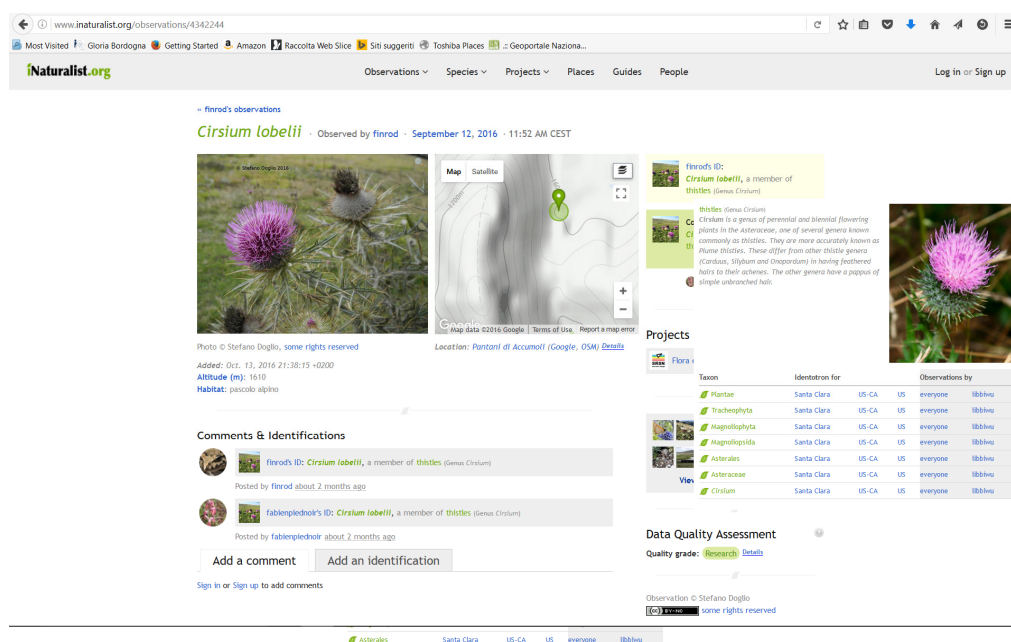

Figure 1. Screenshot of the *inaturalist* Web geoportal displaying a VGI in the form of a classified picture (on the left), its geofootprint with associated imprecision (the circle on the map) and the correspondent visual ontology item with complementary information on the species "Thistle" (on the right) [8].

3. Creating and Managing Contextualized VGI within Space4Agri Project

The Space4Agri (S4A) project defined and developed a Spatial Data Infrastructure (SDI) integrating multisource heterogeneous data from space, i.e., satellite images and Unmanned Aerial Vehicle (UAV) photographs, authoritative databases from Lombardy Agricultural Authority and cadastral data, and in situ data on crop development and agro-practices created by volunteer agronomists, farmers and citizens to support sustainable and precision agriculture in Lombardy region, Northern Italy [9]. A smart application, the S4A smart app, has been designed and implemented to create contextualized VGI by having as cornerstones the following design concepts:

- the possibility to support data normalization and semantic interoperability by providing a domain ontology to the author in order to ease both the creation of observations and the interpretation of contents by potential consumers;
- the reduction and, when possible, resolution of the geometric imprecision of the VGI geofootprint; and
- the sharing of VGI by standard Web services recommended by the Open Geospatial Consortium (OGC) so as to enable its discovery and access with filtering options.

In the following subsections, we recall the main characteristics of the S4A smart app and the workflow of VGI creation, management and fruition on the Web. Finally, we recap the lessons learned based on the experiences of volunteers who used the S4A smart app during the agronomic season of 2015 and 2016 to create in situ data on extensive crops in Lombardy region (Italy) and in Provence (south of France).

3.1. VGI Creation

The S4A smart app (freely downloadable from Google Play Store) can be installed on an Android device by which volunteers can create VGI of agronomic interest, supported by a compact hierarchical domain ontology, a taxonomy with three levels.
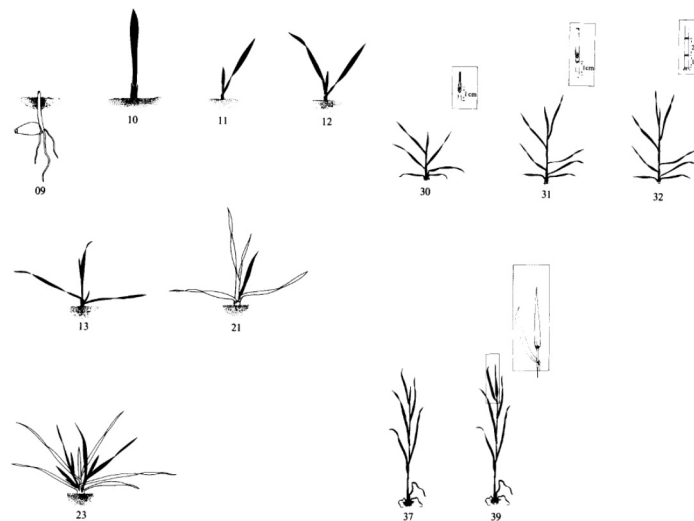
Depending on the context for which the smart app is customized, i.e., agronomic parcels in a region of interest, volunteers are guided to tag entities by selecting categories (from the domain ontology) describing some interesting features or classes of the observed entities, which are visualized on the mobile device screen by a hierarchical menu. This creation modality serves the twofold objective of normalizing VGI content by constraining the volunteers towards both a correct syntactic creation and a consistent semantic annotation of the free text and/or photograph that can also be provided for enriching the description of their observations. In fact, free text can be a valuable means to describe unexpected or unusual situations that could not be pointed out simply by means of a tag selected from an ontology. VGI authors can use free text to express their doubts on the observations, or to describe the conditions that may bias their observations, for example that they were far from the object. Pictures are also a useful mean to document an observation more objectively, since they allow VGI users to directly verify the observations. For example, if a volunteer has specified the observation of an unusual crop for a given area, he/she can prove the observation by providing a picture to support it.

Figure 2 reports an extract from the extensive BBCH ontology defined in [10] and used by the agronomists to specify the development stages of crops [11] (Table 1 reports the principal crop stages categories and descriptions while Figure 2 displays an example of the second level categories, referred to cereals growth stages). It can be seen that each code uniquely identifies a picture and an associated textual description. For inexperienced users, it may be questionable to tell apart the differences between some close codes by observing the pictures or by reading their textual descriptions which contain vague terms such as "*30: beginning of stem elongation ...*" and "*31: first node at least one cm ...*".

The S4A smart app stores a compact version of the same ontology on the mobile device. The volunteer then browses the two level menu and selects the BBCH codes whose description best matches the current observation (see Figure 3).

Table 1. Principal growth stages of crops [10].

Stage	Description
0	Germination/sprouting/bud development
1	Leaf development (main shoot)
2	Formation of side shoots/tillering
3	Stem elongation or rosette growth/shoot development (main shoot)
4	Development of harvestable vegetative plant parts of vegetatively propagated organs/booting (main shoot)
5	Inflorescence emergence (main shoot)/heading
6	Flowering (main shoot)
7	Development of fruit
8	Ripening or maturity of fruit and seed

Cereals

© 1989: Bayer

Growth stage	Code	Description
0: Germination	00	Dry seed (caryopsis)
	01	Beginning of seed imbibition
	03	Seed imbibition complete
	05	Radicle emerged from caryopsis
	06	Radicle elongated, root hairs and/or side roots visible
	07	Coleoptile emerged from caryopsis
	09	Emergence: coleoptile penetrates soil surface (cracking stage)
1: Leaf development	10	First leaf through coleoptile
	11	First leaf unfolded
	12	2 leaves unfolded
	13	3 leaves unfolded
	1.	Stages continuous till...
	19	9 or more leaves unfolded
2: Tillering	20	No tillers
	21	Beginning of tillering: first tiller detectable
	22	2 tillers detectable
	23	3 tillers detectable
	2.	Stages continuous till...
	29	End of tillering. Maximum no. of tillers detectable
3: Stem elongation	30	Beginning of stem elongation: pseudostem and tillers erect, first internode begins to elongate, top of inflorescence at least 1 cm above tillering node
	31	First node at least 1 cm above tillering node
	32	Node 2 at least 2 cm above node 1
	33	Node 3 at least 2 cm above node 2
	3.	Stages continuous till...
	37	Flag leaf just visible, still rolled
	39	Flag leaf stage: flag leaf fully unrolled, ligule just visible
4: Booting	41	Early boot stage: flag leaf sheath extending
	43	Mid boot stage: flag leaf sheath just visibly swollen
	45	Late boot stage: flag leaf sheath swollen

Growth stage	Code	Description
	47	Flag leaf sheath opening
	49	First awns visible (in awned forms only)
5: Inflorescence emergence, heading	51	Beginning of heading: tip of inflorescence emerged from sheath, first spikelet just visible
	52	20% of inflorescence emerged
	53	30% of inflorescence emerged
	54	40% of inflorescence emerged
	55	Middle of heading: half of inflorescence emerged
	56	60% of inflorescence emerged
	57	70% of inflorescence emerged
	58	80% of inflorescence emerged
	59	End of heading: inflorescence fully emerged
6: Flowering, anthesis	61	Beginning of flowering: first anthers visible
	65	Full flowering: 50% of anthers mature
	69	End of flowering: all spikelets have completed flowering but some dehydrated anthers may remain
7: Development of fruit	71	Watery ripe: first grains have reached half their final size
	73	Early milk
	75	Medium milk: grain content milky, grains reached final size, still green
	77	Late milk
8: Ripening	83	Early dough
	85	Soft dough: grain content soft but dry. Fingernail impression not held
	87	Hard dough: grain content solid. Fingernail impression held
	89	Fully ripe: grain hard, difficult to divide with thumbnail
9: Senescence	92	Over-ripe: grain very hard, cannot be dented by thumbnail
	93	Grains loosening in day-time
	97	Plant dead and collapsing
	99	Harvested product

Figure 2. BBCH extract for cereals growth stages. Source: [10].

The created VGI items are automatically geo-referenced by associating them with the geographic coordinates of the current position determined by the GPS sensor of the mobile device. In order to reduce uncertainty, the volunteer can either confirm this geo-reference (being aware of its imprecision which is displayed by a circle whose radius is directly proportional to the imprecision and centered in the location detected by the GPS sensor) or can move the pin to a different location just by clicking on the visualized map. This feature is a useful option that the volunteer can exploit either to correct imprecise geo-localization supplied by the GPS sensor, thus increasing the positional accuracy of the created VGI, or to set the geo-reference of VGI close to the position of the observed entities when they are at a sight view from the current position.
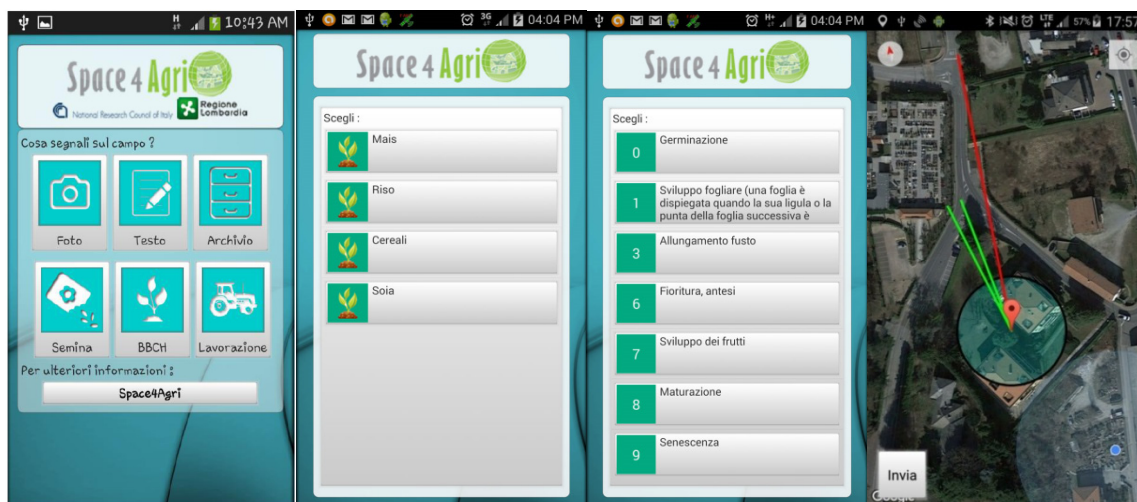

Figure 3. S4A smart app screenshots showing the main menu for VGI creation, the first and second level entries of the compact BBCH agronomic ontology and the representation of the GPS imprecision and camera orientation (depicted by the green line angle).

3.2. VGI Management

The S4A smart app communicates via wireless or mobile network with a back-end application installed and running on a Web server where a geodatabase stores the VGI items. The back-end application can send/receive JavaScript Object Notation (JSON) messages to/from the S4A smart app, and can read/write into the geodatabase. In this phase, the reduction and resolution of the imprecision of the VGI geofootprints is performed by applying a knowledge-based conflation. When receiving a new VGI item, the back-end application analyzes its geofootprint and possibly conflates it with those of other VGI items already stored in the geodatabase, by exploiting the "conflation data layer" (when provided). This represents the known geographic entities of interest for the current project which have to be considered as the targets of volunteers observations. In S4A project, the entities of interest are cultivated fields and, thus, the S4A smart app back-end associates the agronomic categories of the BBCH ontology specified in VGI items to the agronomic parcels surveyed and stored in the agronomic cadastral database, which is provided as the conflation data layer. VGI items outside the fields are dealt with as outliers and stored in a different layer (Figure 4 shows some fields tagged with VGI). The administrator can change the conflation data layer to customize the framework to different projects and needs: all VGI items with geofootprint inside the boundary of an entity of the conflation data layer are associated with that specific entity. In the first version of the S4A smart app, the VGI geofootprints which were far from any entity, were associated with a null entity. Based on user feedback, we developed a second version of the app that associates these VGI items with more entities to distinct degrees computed by exploiting knowledge of both the imprecision of the GPS localization and camera orientation, detected when creating the VGI items. This contextual information

is associated with the VGI item when it is created and is stored in the geodatabase together with VGI geofootprint: specifically we store the associated GPS imprecision radius and the azimuth angle of the camera orientation detected when the volunteer has taken the picture, and thus when the camera frames the target parcel. The degree of association of a VGI item to a cadastral parcel can be computed inversely proportional to the distance of the VGI geofootprint that intersects the parcel boundary in the direction of the camera. By storing contextualized VGI we enable the user exploration of the uncertainty and imprecision of the VGI geofootprints. The "conflation data layer" makes sense when the footprints of the entities of interest are known, such as in our case or in cases of projects collecting reviews on features like Points of Interest. Nevertheless, in the case the entities of interest do not have a known localization, such as in the case of birds or flowers in *inaturalist* project, the "conflation data layer" could still be useful to group the VGI items that have a geofootprint within the boundary of contextual entities of interest, such as the boundary of parks or natural ecosystems. On the other side, if VGI items are created far from the entities of interest, they can still be analyzed and either validated by a moderator or filtered out as noise. For example, in *inaturalist* VGI of flora and fauna in unexpected ecosystems might be important information on environmental changes.
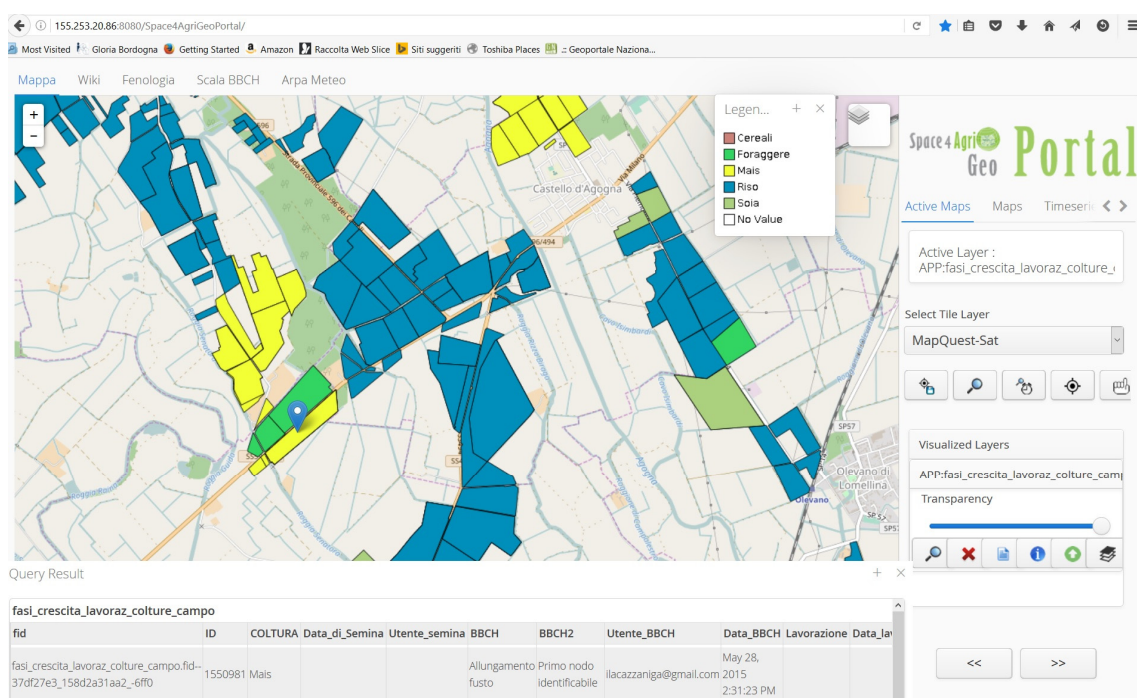

Figure 4. S4A geoportal displaying crop fields tagged with VGI; the distinct colors of the fields represent distinct crop types. The table displays all the VGI items corresponding to the observation of the BBCH growth stage in the cornfield marked with the blue pin.

A Web GIS server deploys the VGI items stored into the geodatabase as distinct views on the Web, together with the conflation data layer and, possibly, other multisource geospatial data of interest to the project via OGC Web Services. A metadata record, based on INSPIRE and the Italian National regulation "Repertorio Nazionale dei Dati Territoriali", is created for each type of VGI layer with basic information describing the common characteristics of its VGI items, i.e., category type, link to the URL of the used extensive BBCH ontology, common keywords to empower data discovery, and the bounding box containing all VGI geofootprints. This way the users of VGI have all extensive information on the semantic and can thus interpret correctly the meaning of VGI. To this end, the S4A geoportal allows accessing the created VGI together with other geo-data of the project. The geoportal can display the distinct views of the VGI items: as pins that one can click and open to see the enclosed

pictures and free text, and as markers with distinct colors to distinguish the types of crops of the BBCH ontology described in the associated legend. One can also see distinct colors in the entities of the conflation data layer that have been tagged with a BBCH category by at least a VGI item (see Figure 4) and, finally, the map of VGI outliers not associated with any surveyed field.

3.3. Lessons Learned within the Space4Agri Project

During the vegetation season of 2015 and 2016, the S4A smart app has been used by farmers, agronomists, students of an agronomic high school and S4A researchers to collect and document the growth of crops within different study areas of Lombardy Region. The S4A smart app has also been used during a campaign in Provence region (south France) to take pictures and annotate lavender fields. Table 2 shows some information about the VGI reports and engaged volunteers.

Table 2. Summary of the number of volunteers (by category) and VGI items (by category) created within S4A project during 2015.

a	Number of Volunteers that installed the S4A smart app	85	100%
b	Number and percentage of active volunteers (who created at least 10 VGI items)	21	25%
c	Number and percentage of farmers/agronomers	28	33%
d	Number and percentage of VGI items about BBCH stages	2594	88%
e	Number and percentage of VGI items as free text and possibly pictures	370	13%
f	Number and percentage of VGI items about agro-practices	137	5%
g	Number and percentage of VGI items about seeding	223	8%
h	TOTAL number of VGI items ($d + f + g$)	2954	100%

It can be seen that active volunteers are 25% of those who installed the S4A smart app.

Most of the VGI items (88%) have been created based on the BBCH ontology, while 5% have been created by the support of the agro-practice ontology. Finally, 13% of VGI items are in the form of semantically annotated text, i.e., they are VGI items specified both by a BBCH code or agro-practice code and free text or picture.

By analyzing the semantically annotated VGI items, we noticed that the texts or pictures were created to report a difficulty or doubt of the volunteer when selecting the BBCH code for different reasons:

- doubt in interpreting the meaning of the text associated with a BBCH code;
- difficulty in clearly distinguishing the characteristic aspects of a BBCH code in the observed crop sample because of a deficiency of the observation means (far point of view); and
- hesitancy to select a unique BBCH code for several observed crop samples close in space within the same parcel, because of the variability of their characteristics.

This suggested to us the need to extend the ontology based reasoning by representing the uncertainty of the volunteer in creating VGI items when selecting tags from the BBCH ontology.

On the user side, when analyzing the stored VGI by means of the S4A geoportal facilities, users sometimes found it difficult to disambiguate the georeference of VGI, when volunteers did not use the manual repositioning of the geofootprint on the map. Especially during the campaign of 2016, many pictures have been taken by automatically georeferencing them using the position determined by the GPS sensor installed on the mobile phones. This practice resulted in many VGI items located on country roads or paths nearby several agronomic areas.

As depicted in Figure 5, it can be seen that many VGI items, the red points, are located on the paths and not on the agronomic fields. This situation generated ambiguities to the user when he/she had to identify the actual area represented in the picture and demanded some means to cope with it. Based on this user feedback, we introduced the representation of GPS imprecision and camera orientation implemented in the second version of the S4A smart app (see the rightmost panel in Figure 3).
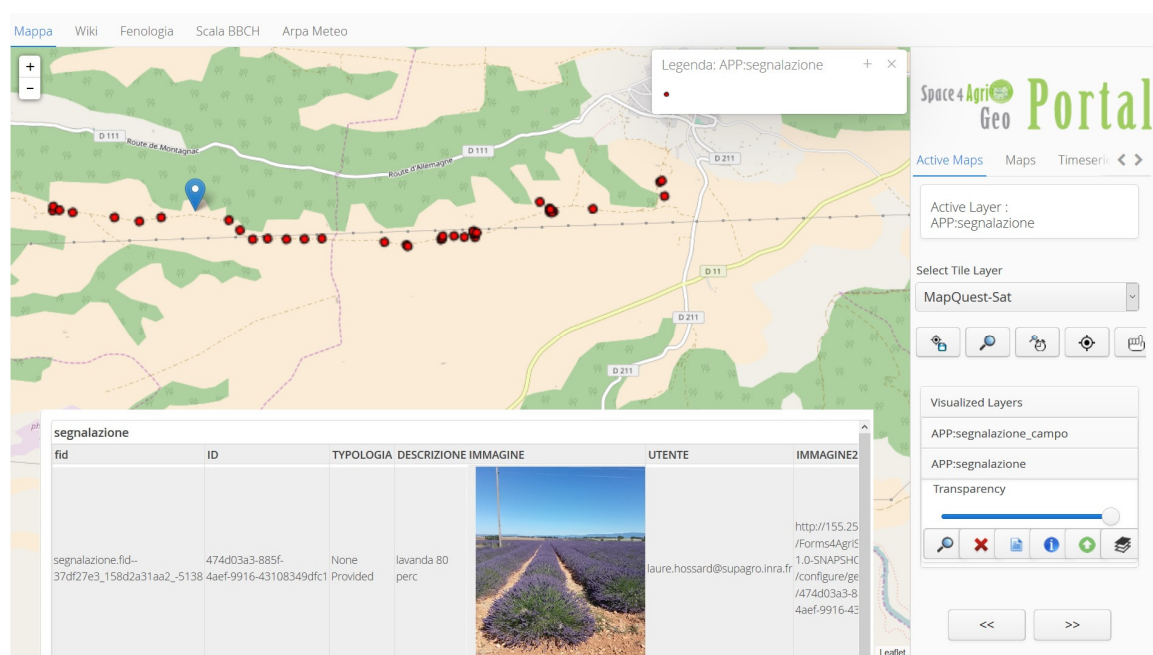
Figure 5. S4A geoportal displaying VGI items (red points) containing pictures taken from fields in Provence, France, during summer 2016 campaign. The pictures are mostly located on the country roads crossing an agronomic area cultivated with lavender and thus one does not know which parcel they annotate.

4. Representing and Managing Ill-Defined Knowledge by Fuzzy Ontologies

A common way to define ontologies is by means of the Web Ontology Language (OWL) [17,18]. In OWL, the classes are the intensional definitions of the concepts, the relationships between classes are labeled by their properties, and the instances of the classes are the extensional definitions, i.e., the single individuals belonging to the classes. OWL supports the reasoning power of Description Logic (DL). Thus, OWL-DLs can be viewed as machine-processable standardization of Description Logics suitable for interoperability and scalability of systems, which promotes reuse of data and reasoning over the Web. Since OWL-DL allows representing the world in terms of crisp concepts (sets) and relationships among entities (that are either true or false) it becomes unsuitable in domains in which concepts or relationships have not a precise definition but are ill-defined, i.e., vague by their nature. For instance, if we would like to define an ontology about flowers, we may find difficulties in encoding into OWL-DL the ill-defined knowledge about Calla expressed as follows:

$$\text{``Calla is a very large, long white flower on thick stalks''.} \tag{1}$$

The vague concepts like "*very large*", "*long*" and "*thick*" involve some fuzziness for which a crisp and precise definition does not make sense. What is the wideness of a flower that makes it appearing "*very large*"? This is a matter of degrees depending on a subjective interpretation and, certainly, there is not a crisp transition between a Calla being "*very large*" and "*not very large*" shared by all authors.

Another dual situation that may happen in the real world of observations is when the VGI author is not completely sure about his/her observation, either because he/she does not have adequate knowledge of the problem, or because of deficiencies of the means of observation. This may happen even when the domain knowledge is encoded into a precise ontology.

Finally, there are situations that may involve both ill-defined knowledge and uncertainty of VGI authors. Let us consider this description of sparrows provided by Wikipedia (https://en.wikipedia.org/wiki/Sparrow):

"sparrows are plump **little** brown or **greyish** birds, **often** with black, yellow or white markings. **Typically** 10–20 cm long, they range in size from the chestnut sparrow (Passer eminibey), at 11.4 cm and 13.4 g, to the parrot-billed sparrow (Passer gongonensis), at 18 cm and 42 g. They have **strong**, **stubby** conical beaks with **decurved** culmens and **blunter** tips".

This taxonomy of sparrow types contains both fuzzy (identified by the adjectives in bold font) and precise predicates (measures are provided in centimeters and gram). As far as fuzzy predicates, it might be questionable to state if a bird is little and greyish. Besides, it might even be impossible to measure the actual length and weight of an observed bird or to estimate the measurements from a far observation point; thus, one might find questionable to tag the observation as that of either a chestnut sparrow or a parrot-billed sparrow, while one could express the uncertainty or doubt about the truth of the precise predicates.

Fuzzy ontologies have been defined in order to represent the fuzziness of concepts in an ontology. According to [14], a fuzzy ontology can be defined by an extension of OWL-DL, specifically Fuzzy OWL-DL. In order to introduce the basic concepts, we first need to recall the definition of a fuzzy set. Fuzzy sets were introduced by Zadeh in 1965 [24] to deal with vague concepts.

Formally, a fuzzy set A with respect to a universe X has a membership function $\mu_A: X \rightarrow [0, 1]$, assigning a membership degree, $\mu_A(x) \in [0, 1]$, to each element x of the domain X.

Typically, if $\mu_A(x) = 1$ it means that x definitely belongs to A, while, if $\mu_A(x) = 0$ it means that x does not belong to A at all. $\mu_A(x) = 0.8$ means that x is partially an element of A, which could mean either that x does not satisfy all properties that characterize an element of A or that one lacks complete knowledge on x and cannot state precisely if it is an element of A (one must choose one of the two possible interpretations and both cannot be assumed at the same time).

Accordingly, in fuzzy logics, the notion of degree of membership, $\mu_A(x)$, of an element $x \in X$ to the fuzzy set A over X is regarded as the degree of truth in [0, 1] of the statement "x is A". This interpretation is applied in fuzzy DL, where a concept A, rather than being interpreted as a classical set, is interpreted as a fuzzy set and, thus, concepts can be fuzzy. As a consequence, the statement "x is A", i.e., x:A, has a degree of truth in [0, 1] given by $\mu_A(x)$ the degree of membership of the individual element x to the fuzzy set A. Boolean operators defined to combine classic sets have been extended to combine fuzzy sets so that the logic intersection, union and complement are respectively defined by a t-norm (min), a t-conorm (max), and not (1-) operators.

By means of Fuzzy OWL-DL, one can represent an ill-defined statement like definition (1) describing a Calla by the following axiom with fuzzy predicates:

$$\text{Calla} \subseteq \text{Flower} \cap (\exists\text{hasSize.very(Large)}) \cap (\exists\, \text{hasPetalWidth.Long}) \cap \\ (\exists\text{hasColour.White}) \cap \exists\text{hasStalks.Thick} \tag{2}$$

where

- Flower is a crisp predicate so that $\mu_{\text{Flower}}(x) = 1$ if x is a flower, otherwise $\mu_{\text{Flower}}(x) = 0$. Notice that crisp set are a particular case of fuzzy sets.
- Size.very_Large, PetalWidth.Long, Colour.White and Stalks.Thick are fuzzy predicates represented by fuzzy sets with membership functions $\mu_{\text{Size.very\_Large}}$, $\mu_{\text{PetalWidth.Long}}$, $\mu_{\text{Colour.White}}$, and $\mu_{\text{Stalks.Thick}}$, as depicted in Figure 6.
- $\cap$ is the intersection operator between fuzzy sets defined by the minimum, and $\subseteq$ is the subset inclusion operator.

Let us assume that the VGI author can provide a precise observation of a flower x by measuring the size, the petal width, and the type of stalk. Given these precise measurements and the membership functions $\mu_{\text{Size.very\_Large}}$, $\mu_{\text{PetalWidth.Long}}$, $\mu_{\text{Colour.White}}$, and $\mu_{\text{Stalks.Thick}}$ defining the meaning of the respective fuzzy predicates, we can compute their degrees of satisfaction and, finally, the degree of

truth ($\mu_{Calla}(x)$) of the statement "x *is a Calla*" by applying approximate reasoning based on fuzzy predicates [17]:

$$\mu_{Calla}(x) = \min(1, \mu_{Size. Large}(x), \mu_{PetalWidth.Long}(x), \mu_{Colour.White}(x), \mu_{Stalks.Thick} (x)) \qquad (3)$$
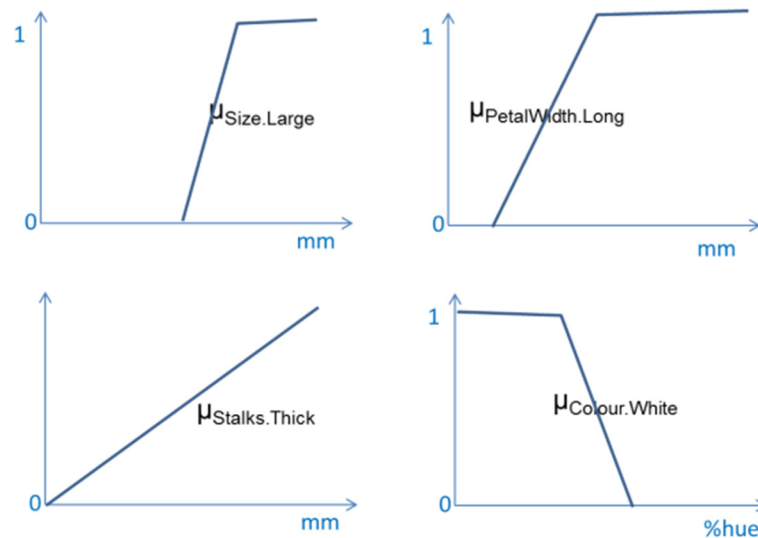

Figure 6. Membership functions of the fuzzy sets **Size.large**, **PetalWidth.Long**, **Stalks.Thick**, and **Colour.White**.

Nevertheless, the approximate reasoning based on a fuzzy ontology illustrated so far is suitable to deal with ill-defined knowledge, i.e., fuzzy knowledge, while we need to reason with precise knowledge and uncertain observations, as outlined by the lessons learned in Section 3.3.

In fact, there is often a misunderstanding on the modeling of uncertainty and imprecision or fuzziness which are wrongly considered synonymous terms [25]. Nevertheless, under uncertainty theory, statements are either completely true or completely false in any world/interpretation but we do not know which world is the right one, so we define a probability or possibility distribution over the worlds. For example, the statement "*x* is a Wheat crop" is a crisp one: *x* can be either a Wheat crop or not, it cannot be partially a Wheat crop, the degree that we can associate to this statement is relative to our knowledge on the truth about *x* being a Wheat crop, which may depend on some deficiency of the observation. In fuzzy theory, statements are true and false to a degree since they contain fuzzy concepts such as in a fuzzy ontology: for example the statement "*x* has a Long leaf", contains the fuzzy term Long whose meaning in cm is a matter of degree, the longer the leaf the greater the degree by which the statement is satisfied. We can model observations affected by some deficiency of the conditions by two alternative statements [26]: for example, by observing a wheat crop from a far point of view one could either specify his/her uncertainty on the truth of a precise statement such as "*I am 0.6 certain that the leaf length is 40 cm*" or one could express a certain fuzzy statement such as "*I am sure that x has a Long leaf*" or both an uncertain and fuzzy statement such as "*I am 0.8 certain that the leaf length is between 35 and 45 cm*". In these statements, we can notice that the uncertainty degrees are inversely related with the amount of imprecision/fuzziness of the predicates. One can guess that the total amounts of uncertainty plus imprecision/fuzziness in all statements describing the same observation in alternative ways are constant and depend on the degree of *overall defect* of the observation, and the greater the defect of the observation the greater is the total amount of uncertainty plus imprecision/fuzziness of the statement that describes the observation. In the following section we will describe our proposal to cope with imprecisions and uncertainty of observations when creating VGI by introducing uncertainty level-based approximate reasoning with a fuzzy ontology.

5. Coping with Uncertainty and Imprecision of Observations Using a Fuzzy Ontology with Uncertainty Level-Based Approximate Reasoning

Hereafter, we will make an example relative to the definition of a fuzzy BBCH ontology describing the crop types, considered as precise concepts, and their development stages introduced in Section 3.1, that can be considered as fuzzy stages to some extent since there is a gradual transition from one development stage to the successive one with intermediate characteristics of the crop growth. Thus, it makes sense to define fuzzy BBCH values with membership functions on a continuous domain of the ordered crisp BBCH stages defined in Table 1 (see Figure 7).

In order to cope with the deficiency of observations, we will allow volunteers to express the *overall defect* (uncertainty + imprecision/fuzziness) of their observation by a numeric value d in [0, 1] on both the observed crop type and relative development stage, where the development stage can be specified by selecting from the fuzzy ontology a fuzzy BBCH values.

Since crop types are considered precise concepts, indeed the overall defect degree d reduces to an uncertainty degree. To cope with the volunteer's uncertainty in selecting a crop type we allow VGI authors to select more than a single crop type for each observed instance and we allow associating with each selected crop an indication of his/her uncertainty on the observation. In principle, we could allow volunteers to specify more crop types with distinct uncertainty degrees d_i in [0, 1], such that $\Sigma_{i = 1, \dots, n}\, d_i = 1$, where 0 means full certainty, 1 means full uncertainty and intermediate values are indications of partial uncertainty.

To simplify the user interaction when using the S4A smart app we do not ask volunteers to specify a numeric value d in [0, 1], which would require the burden task of digitizing numbers, but we limit the possible choices to three values $d \in \{0, 0.5, 1\}$, that they can select by simply clicking on the red, yellow or green button of a traffic light displayed on the screen of the smart app, which seems a more intuitive and faster interaction. The model is general and would also work in the case the volunteer is permitted to specify any d in [0, 1]. Thus, the degree can be chosen in $\{0, 0.5, 1\}$ where 0 means certain, 0.5 means a little uncertain, and 1 means uncertain.

By exploiting the order existing between the developing stages, we define fuzzy BBCH values with triangular membership functions on the crisp BBCH scale as depicted in Figure 7. The domain is considered as continuous since there is a gradual transition between development stages. Each membership function can be defined by a triple ($a \leq b \leq c$) where a, b, $c \in \{0, ..., 9\}$ are ordered crisp BBCH stages as follows:

$$\mu_{FuzzyBBCH}(x) = \begin{cases} 0 & if & x \leq a \\ \frac{b-x}{b-a} & if & a < x < b \\ \frac{c-x}{c-b} & if & b < x < c \\ 0 & if & x \geq c \end{cases} \qquad (4)$$

in which x identifies an observation. This choice was done based on the idea that, in principle, we allow volunteers selecting fuzzy BBCH values from the fuzzy ontology and the overall defect degree d can be considered as comprising both uncertainty and imprecision of the observation: $d = 0$ means certain and precise observation, then the specified fuzzy BBCH value must be interpreted as a precise observation, i.e., a crisp BBCH stage; $d = 1$ means maximum uncertain and imprecise observation, then the specified fuzzy BBCH value must be interpreted with maximum fuzziness, i.e., more BBCH stages are possible; by decreasing d towards 0 it means that the observation becomes less uncertain and less imprecise/fuzzy, and thus the selected fuzzy BBCH value must be interpreted as less fuzzy, i.e., the actual BBCH stages decrease with d until only one is possible. To model this behavior, during the reasoning phase, we interpret d as a threshold on the membership function of the specified fuzzy BBCH so that only the crisp BBCH stages with degrees equal or greater than $(1 - d)$ are considered as the possible values of the development stage. The triangular shaped membership functions have a simple shape with a point-like nucleus, i.e., only one single value of their domain (BBCH stages) has

a full membership degree. This is the reason we adopted triangular shaped functions with respect to trapezoidal ones so that, in the case of certain and precise observation, ($d = 0$), only one precise BBCH stage is possible:

- $d = 1$ means full uncertainty and maximum fuzziness of the observation. As an example, Figure 7 shows the membership functions of the fuzzy BBCH values defined on the domain of the crisp BBCH stages of Table 1. Assume the defect $d = 1$ in observing the development stage ~Flowering. Being d the complement of a threshold on the membership function of ~Flowering the crisp BBCH stages (i.e., the values on the domain of the membership function) that have membership degrees above 0 are only the following ones: Flowering is fully possible (degree 1), the stages Heading and Fruit Development are partially possible (having both the degree 0.75), and Booting and Ripening are also minimally possible (having both the degree 0.25). All the other BBCH stages have degree 0 with respect to the membership function of ~Flowering and thus are excluded as possible development stages of the faulty observation.

- when $d = 0.5$, it means some uncertainty and partial fuzziness of the observation. In this case, due to the effect of the decreased d with respect to the previous situation we can reduce the fuzziness of the observation: in this case, the threshold d allows excluding both Booting and Ripening as possible stages (since their degree being 0.25 is below the threshold 0.5), while Flowering is still fully possible (degree 1 is above 0.5) and Heading and Fruit Development are still partially possible (their degree 0.75 is also above 0.5). With respect to the previous worse observation of ~Flowering, that resulted with five possible development stages, now we have only three possible development stages.

- when $d = 0$ it means certain and precise observation. In this case only one crisp BBCH stage is fully possible (Flowering) since this is the only development stage whose membership degree to ~Flowering is equal to 1, while all the other development stages of the domain can be excluded as possible stages of the observation. In case of perfect observation, even if we selected a fuzzy BBCH value, we resulted with only one possible precise BBCH stage.
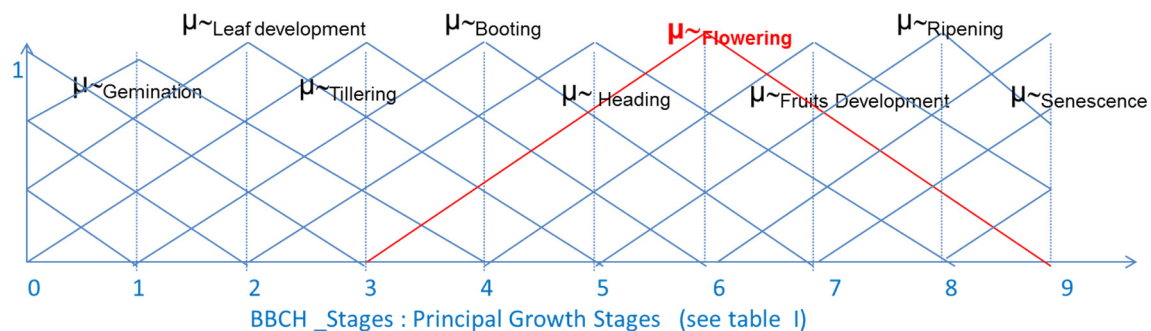

Figure 7. membership functions of the Fuzzy BBCH values (identified by ~prefix) defined on the BBCH stages of the first level in Table 1.

Formally, given a faulty observation x of a fuzzy BBCH value (~*BBCH_Value*) with defect d, it means we are at most $(1 - d)$ certain on the truth of the fuzzy predicate "the development stage of x is ~*BBCH_Value*" (i.e., *Certain* (x is ~*BBCH_Value*) $\leq 1 - d$). Based on this premise, and on the membership function of ~*BBCH_Value* we can compute the degree of possibility that "x is $BBCH_{Stage}$"

is true; that is, the possibility that the crisp BBCH stage ($BBCH_{Stage}$) is a possible development stage of *x* as follows:

$$
\begin{aligned}
&\text{Given} \quad Certainty \left(x \text{ is } \sim BBCH_{Value} \right) \leq 1 - d \quad \text{then} \\
&Possibility \left(x \text{ is } BBCH_{Stage} \right) = \mu_{BBCH\ Stage} \left(x \right) \quad \text{where} \\
&\mu_{BBCH\ Stage} \left(x \right) = \begin{cases} \mu_{\sim BBCH_{Value}} \left(BBCH_{Stage} \right) & \text{if } \mu_{\sim BBCH_{Value}} \left(BBCH_{Stage} \right) \geq 1 - d \\ 0 & \text{if } \mu_{\sim BBCH_{Value}} \left(BBCH_{Stage} \right) < 1 - d \end{cases}
\end{aligned} \tag{5}
$$

Let us make a simple example of uncertainty level-based approximate reasoning applying the previous formula.

Assume the following ten axioms to be an extract of the fuzzy TBOX ontology, where each predicate admits a degree of possibility in [0, 1] to be true, and let us define the intersection and inclusion by the min:

1. Flowering_Wheat $\subseteq$ crop $\cap$ ($\exists$Wheat) $\cap$ ($\exists$Flowering)
2. Heading_Wheat $\subseteq$ crop $\cap$ ($\exists$Wheat) $\cap$ ($\exists$Heading)
3. Fruit_Development_Wheat $\subseteq$ crop $\cap$ ($\exists$Wheat) $\cap$ ($\exists$Fruit_Development)
4. Booting_Wheat $\subseteq$ crop $\cap$ ($\exists$Wheat) $\cap$ ($\exists$Booting)
5. Ripening_Wheat $\subseteq$ crop $\cap$ ($\exists$Wheat) $\cap$ ($\exists$Ripening) $\qquad$ (6)
6. Flowering_Barley $\subseteq$ crop $\cap$ ($\exists$Barley) $\cap$ ($\exists$Flowering)
7. Heading_Barley $\subseteq$ crop $\cap$ ($\exists$Barley) $\cap$ ($\exists$Heading)
8. Fruit_Development_Barley $\subseteq$ crop $\cap$ ($\exists$Barley) $\cap$ ($\exists$Fruit_Development)
9. Booting_Barley $\subseteq$ crop $\cap$ ($\exists$Barley) $\cap$ ($\exists$Booting)
10. Ripening_Wheat $\subseteq$ crop $\cap$ ($\exists$Wheat) $\cap$ ($\exists$Ripening)

Just to clarify the meaning, the first axiom states that a wheat crop at the flowering phase is a crop, whose type is wheat and whose development stage is Flowering.

Let us assume, as example of ABOX, the following faulty observation performed by a volunteer:

$$x \subseteq (\text{Wheat}, d_1 = 0) \cap (\sim \text{Flowering}, d_2 = 0.5)$$

This means that the volunteer is certain that *x* is a Wheat crop (*Certainty* (Wheat) = 1) while he/she believes that the development stage of *x* is ~Flowering with at most partial certainty (*Certainty* (~Flowering) $\leq$ 0.5). By applying uncertainty level-based approximate reasoning, (Equation (5)), given the degree of maximum certainty of ~Flowering, we can deduce the possible development stages of *x*. This is achieved by evaluating the degree of possibility that the axioms defined in Equation (6) are true as follows:

$$\mu_{\text{FloweringWheat}}(x) = \min(\mu_{\text{Crop}} (x), \mu_{\text{Wheat}}(x), \mu_{\text{Flowering}} (x)) = \min(1, 1, 1) = 1$$

$$\mu_{\text{HeadingWheat}}(x) = \min(\mu_{\text{Crop}} (x), \mu_{\text{Wheat}}(x), \mu_{\text{Heading}} (x)) = \min(1, 1, 0.75) = 0.75$$

$$\mu_{\text{Fruit DevelopmentWheat}}(x) = \min(\mu_{\text{Crop}} (x), \mu_{\text{Wheat}}(x), \mu_{\text{Fruit Development}} (x)) = \min (1, 1, 0.75) = 0.75$$

$$\mu_{\text{BootingWheat}}(x) = \min(\mu_{\text{Crop}} (x), \mu_{\text{Wheat}}(x), \mu_{\text{Booting}}(x)) = \min (1, 1, 0) = 0$$

$$\mu_{\text{RipeningWheat}}(x) = \min(\mu_{\text{Crop}} (x), \mu_{\text{Wheat}}(x), \mu_{\text{Ripening}}(x)) = \min (1, 1, 0) = 0$$

$$\mu_{\text{FloweringBarley}}(x) = \mu_{\text{HeadingCorn}}(x) = \mu_{\text{Fruit DevelopmentBarley}}(x) = \mu_{\text{BootingBarley}}(x) = 0 \text{ since } \mu_{\text{Barley}}(x) = 0$$

A linguistic summary of the observation is the following: "it is *fully possible* that wheat crop *x* is at the flowering stage, but is also *highly possible* that *x* is at the Fruit development or Heading stages, while all other stages can be certainly excluded".

In case the volunteer has specified:

$$x \subseteq (\text{Wheat}, 0.5) \cup (\text{Barley}, 0.5) \cap (\sim\!\text{Flowering}, 0.5)$$

which means that *x* can be either Wheat (with certainty 0.5, i.e., *Certainty*(Wheat) = 0.5) or Barley (with certainty 0.5, i.e., *Certainty*(Barley) = 0.5), and *Certainty*(~Flowering) ≤ 0.5. We would get as a result of the uncertainty level-based approximate reasoning the following possibility degrees for the axioms defined in (6) to be true:

$$\mu_{\text{FloweringWheat}}(x) = \mu_{\text{HeadingWheat}}(x) = \mu_{\text{Fruit DevelopmentWheat}}(x) = 0.5$$

$$\mu_{\text{FloweringBarley}}(x) = \mu_{\text{HeadingBarley}}(x) = \mu_{\text{Fruit Development Barley}}(x) = 0.5$$

This result can be expressed linguistically as follows: "*x* could be either a Wheat or a Barley crop at one of the following stages: either Flowering, or Heading, or Fruit Development, while all the other stages can be excluded".

Within this framework, when *d* = 0 for all observations (they are all fully certain), we obtain the classic reasoning of OWL-DL (the fuzzy selections are indeed crisp, precise observations).

Let us make an example given the following certain observation:

$$x \subseteq (\text{Wheat}, 0) \cap (\sim\!\text{Flowering}, 0)$$

that is, *Certainty*(Wheat) = 1 and *Certainty*(~Flowering) ≤ 1.

In this case, by applying both Equation (5) and the axioms defined in Equation (6), we obtain the precise conclusion:

$\mu_{\text{FloweringWheat}}(x) = 1$, which translates linguistically as follows:

"It is certain that the Wheat crop *x* is at the Flowering stage"

This certainty can be deduced by the fact that Flowering is the only fully possible stage of *x*.

On the other side, when *d* = 1 (worst situation), we are in the framework of approximate reasoning with a fuzzy ontology.

6. Conclusions

In this paper we analyzed several aspects of uncertainty affecting VGI in citizens science projects, with a special focus to projects asking volunteers to classify observations and measurements. More closely, we illustrated the case study example of Space4Agri project [9] and discussed the use of a classic domain ontology and its limitations.

It is undoubtedly true that by constraining volunteers to select labels from a predefined menu limits the potential expressiveness of the volunteers' textual descriptions of real observations which rarely match a prototype perfectly, as outlined by the on-field experimentations carried out within the S4A project. Nevertheless, to ease the VGI processing and interpretation it is an advantage to be able to exploit semantic annotations based on ontology, possibly complemented by free text. To cope with the stiffness that adhering to a classic ontology may require to volunteers, fuzzy OWL ontologies were defined that allow representing ill-defined knowledge and thus modeling a level of flexibility in VGI creation based on them.

We introduce in this paper a further level of flexibility by allowing to model the uncertainty of volunteers when using a fuzzy OWL ontology to tag observations, through uncertainty level-based approximate reasoning. In principle, the adoption of this fuzzy ontology offers several potential advantages with respect to the crisp definition.

- We can model both precise and uncertain creation of VGI, so coping with the limitations of the observation means and context.

- We support inexperienced volunteers who are unable to interpret the meaning of the BBCH stages of crops and to recognize them on the field by allowing them to select "fuzzy" BBCH values; in doing so we can cope also with variability of the observed entities. For example, in the case one observes variable growth stages of the crops within the same parcel he/she can specify a fuzzy BBCH value with a partial defect degree d, thus summarizing with a unique choice several growth stages.

- The VGI consumers, when analyzing the VGI collection created based on fuzzy ontologies, can express flexible queries constraining the VGI items' defects to be below a desired value, so as to retrieve ranked VGI items as in fuzzy databases [27]. They can select only the crop observations having at least a minimum level of possibility to be the true development stages of the observed crops, as computed by applying uncertainty level-based approximate reasoning.

To definitely assess the advantages/drawbacks of the approach we need to carry out a user evaluation. We are now implementing a new version of the S4A smart app using the fuzzy BBCH scale and allowing users to specify the amount of defect of their observations. This is the prerequisite for carrying out a real on field experimentation aimed at evaluating the benefits of using the fuzzy OWL ontologies coupled with uncertainty level-based approximate reasoning.

Author Contributions: Gloria Bordogna wrote the paper and the other co-authors revised it. Moreover, she defined the model to deal with uncertainty and imprecision when creating VGI in S4A project. Luca Frigerio developed both the smart app and back-end and the project personalized geoportal. Tomas Kliment designed the geodatabase and the Web GIS with Web services. Pietro Alessandro Brivio, Laure Hossard and Giacinto Manfron tested and provided the feedback for the revision of the S4A smart app. Simone Sterlacchini was the main tester of the S4A geoportal.

Abbreviations

The following abbreviations are used in this manuscript:

App	Software Application for Smart mobile devices connected to the Internet.
BBCH scale	the Biologische Bundesanstalt, Bundessortenamt and CHemical industry scale is an ontology for a uniform coding of phenologically similar growth stages of all mono- and dicotyledonous plant species commonly used by agronomists [14].
DEM	Digital Elevation Model is a digital model, in the form of either a gridded array or a vector data file, representing surface elevation information.
GPS	the Global Positioning System is a global navigation satellite system that provides location and time information anywhere on or near the Earth where there is an unobstructed line of sight to four or more GPS satellites.
CNR-IDPA	Institute for the Dynamics of Environmental Processes of the National research Council of Italy.
CNR-IREA	Institute for the Electromagnetic Sensing of the Environment of the National research Council of Italy.
OGC	Open Geospatial Consortium.
OWL	Ontology Web Language used to encode ontologies.
OWL DL	OWL DL is so named due to its correspondence with Description Logics, a field of research that has studied the logics that form the formal foundation of OWL.
POI	Point Of Interest identifying a gelocated entity of the real world.
SDI	Spatial Data Infrastructure.
S4A	abbreviation of Space4Agri acronym for the project entitled "Sviluppo di metodologie aerospaziali innovative di osservazione della terra a supporto del settore Agricolo in Lombardia"carried out during 2013–2015 at both CNR IREA and CNR IDPA, co-funded Regione Lombardia and CNR.
UAV	Unmanned Aerial Veicle, also known as "drone", is an aircraft without any human people onboard that is driven by remote.
VGI	Volunteered Geographic Information.
W3C	abbreviation of the World Wide Web Consortium, the international community that develops open standards to ensure the long-term growth of the Web (https://www.w3.org/).

References

1. Haklay, M. Citizen Science and Volunteered Geographic Information—Overview and typology of participation. In *Volunteered Geographic Information, Public Participation, and Crowdsourced Production of Geographic Knowledge*; Sui, D.Z., Elwood, S., Goodchild, M.F., Eds.; Springer: Berlin, Germany, 2012; pp. 105–122.

2. Goodchild, M.F. Citizens as Voluntary Sensors: Spatial Data Infrastructure in the World of Web 2.0. *Int. J. Spat. Data Infrastruct. Res.* **2007**, *2*, 24–32.

3. Bordogna, G.; Carrara, P.; Criscuolo, L.; Pepe, M.; Rampini, A. On predicting and improving the quality of Volunteer Geographic Information projects. *Int. J. Digit. Earth* **2014**, *9*, 1–22. [CrossRef]

4. Schade, S.; Tsinaraki, C. *Survey Report: Data Management in Citizen Science Projects*; JRC Technical Report for European Commission; Publications Office of the European Union: Luxembourg, 2016.

5. Crall, A.W.; Newman, G.J.; Stohlgren, T.J.; Holfelder, K.A.; Graham, J.; Waller, D.M. Assessing citizen science data quality: An invasive species case study. *Conserv. Lett.* **2011**, *4*, 433–442. [CrossRef]

6. Al-Bakri, M.; Fairbairn, D. Assessing the accuracy of 'crowdsourced' data and its integration with official spatial data sets. In Proceedings of the Ninth International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences (Accuracy 2010), Leicester, UK, 20–23 July 2010.

7. Camponovo, M.E.; Freundschuh, S.M. Assessing uncertainty in VGI for emergency response. *Cartogr. Geogr. Inf. Sci.* **2014**, *41*, 440–455. [CrossRef]

8. Naturalist. Available online: http://www.inaturalist.org/ (accessed on 8 August 2016).

9. Bordogna, G.; Kliment, K.; Frigerio, L.; Stroppiana, D.; Brivio, P.A.; Crema, A.; Boschetti, M.; Sterlacchini, S. Spatial Data Infrastructure integrating multisource heterogeneous geospatial data and time series: A study case in agriculture. *ISPRS Int. J. Geo-Inf.* **2016**, *5*, 73. [CrossRef]

10. Hack, H.; Bleiholder, H.; Buhr, L.; Meier, U.; Schnock-Fricke, U.; Weber, E.; Witzenberger, A. Einheitliche Codierung der phänologischen Entwicklungsstadien mono-und dikotyler Pflanzen–Erweiterte BBCH-Skala, Allgemein-Nachrichtenbl. *Deut. Pflanzenschutzd.* **1992**, *44*, 265–270.

11. Growth Stages of Mono and Dicotyledonous Plants. Available online: https://www.politicheagricole.it/flex/AppData/WebLive/Agrometeo/MIEPFY800/BBCHengl2001.pdf (accessed on 8 August 2016).

12. OWL 2 Web Ontology Language Primer. W3C Recommendation. 2009. Available online: http://www.w3.org/TR/owl2-primer/ (accessed on 6 December 2016).

13. Hitzler, P.; Krötzsch, M.; Rudolph, S. *Foundations of Semantic Web Technologies*; CRC Press: Boca Raton, FL, USA, 2009.

14. Straccia, U. All about fuzzy description logics and applications. In *Reasoning Web 2015, LNCS 9203*; Faber, W., Paschke, A., Eds.; Springer: Berlin, Germany, 2015; pp. 1–31.

15. Bobillo, F.; Straccia, U. The fuzzy ontology reasoner fuzzyDL. *Knowl. Based Syst.* **2016**, *95*, 12–34. [CrossRef]

16. Bakillah, M.; Liang, S.H.L.; Zipf, A.; Arsanjani, J.J. Semantic interoperability of sensor data with volunteered geographic information: A unified model. *ISPRS Int. J. Geo-Inf.* **2013**, *2*, 766–796. [CrossRef]

17. Gruber, T.R. Toward principles for the design of ontologies used for knowledge sharing. *Int. J. Hum. Comput. Stud.* **1995**, *43*, 907–928. [CrossRef]

18. Cho, W.C.; Richards, D. Ontology construction and concept reuse with formal concept analysis for improved web document retrieval. *Web Intell. Agent Syst.* **2007**, *5*, 109–126.

19. Ronzhin, S. Semantic Enrichment of Volunteered Geographic Information Using Linked Data: A Use Case Scenario for Disaster Management. Master's Thesis, University of Twente, Enschede, The Netherlands, 2015.

20. Stadler, C.; Lehmann, J.; Höffner, K.; Auer, S. LinkedGeoData: A core for a web of spatial open data. *Semant. Web* **2012**, *3*, 333–335.

21. Bordogna, G.; Boschetti, M.; Brivio, P.A.; Crema, A.; Stroppiana, D.; Frigerio, L.; Kliment, T.; Sterlacchini, S. Semantic interoperability of volunteered geographic information based on contextual knowledge. In Proceedings of the Link-VGI Workshop of AGILE 2016, Helsinki, Finland, 14–17 June 2016.

22. Martínez Ramos, I.; Vandecasteele, A.; Devillers, J. Semantic integration of authoritative and Volunteered Geographic Information (VGI) using ontologies. In Proceedings of the 14th International Conference on Agile Software Development, Vienna, Austria, 3–7 June 2013.

23. Pazoky, S.H.; Karimipour, F.; Hakimpour, F. An ontological solution for perceptual uncertainties of VGI. In Proceedings of the Eighth International Conference on Geographic Information Science, Vienna, Austria, 23–26 September 2014.

24. Zadeh, L.A. Fuzzy sets. *Inf. Control* **1965**, *8*, 338–353. [CrossRef]

25. Straccia, U. *Towards a Fuzzy Description Logic For The Semantic Web*; Springer: Berlin, Germany, 2005.

26. Bordogna, G.; Pasi, G. Modeling linguistic qualifiers of uncertainty in a fuzzy database. *Int. J. Intell. Syst.* **2000**, *15*, 995–1014. [CrossRef]

27. Bordogna, G.; Psaila, G. Customizable Flexible Querying in Classical Relational Databases. In *Handbook of Research on Fuzzy Information Processing in Databases*; Galindo, J., Ed.; IGI Global: Hershey, PA, USA, 2008; pp. 191–217.