

A Data Fusion-based Framework for Integrating VGI to Update and Enrich Authoritative LULC Data

Lanfa Liu, Ana-Maria Olteanu-Raimond,
Laurence Jolivet, Arnaud-le Bris

Lanfa Liu, Ana-Maria Olteanu-Raimond, Laurence Jolivet, Arnaud-le Bris & Linda See (2020) A data fusion-based framework to integrate multi-source VGI in an authoritative land use database, *International Journal of Digital Earth*, DOI: 10.1080/17538947.2020.1842524



Lanfa.Liu@outlook.com

Content

- Introduction
- Data Description
- Dempster-Shafer Theory (DST)
- Results and Discussion
- Conclusion

- Earth Observation Data ->LULC data
 - CORINE Land Cover (CLC), Urban Atlas (UA), etc.



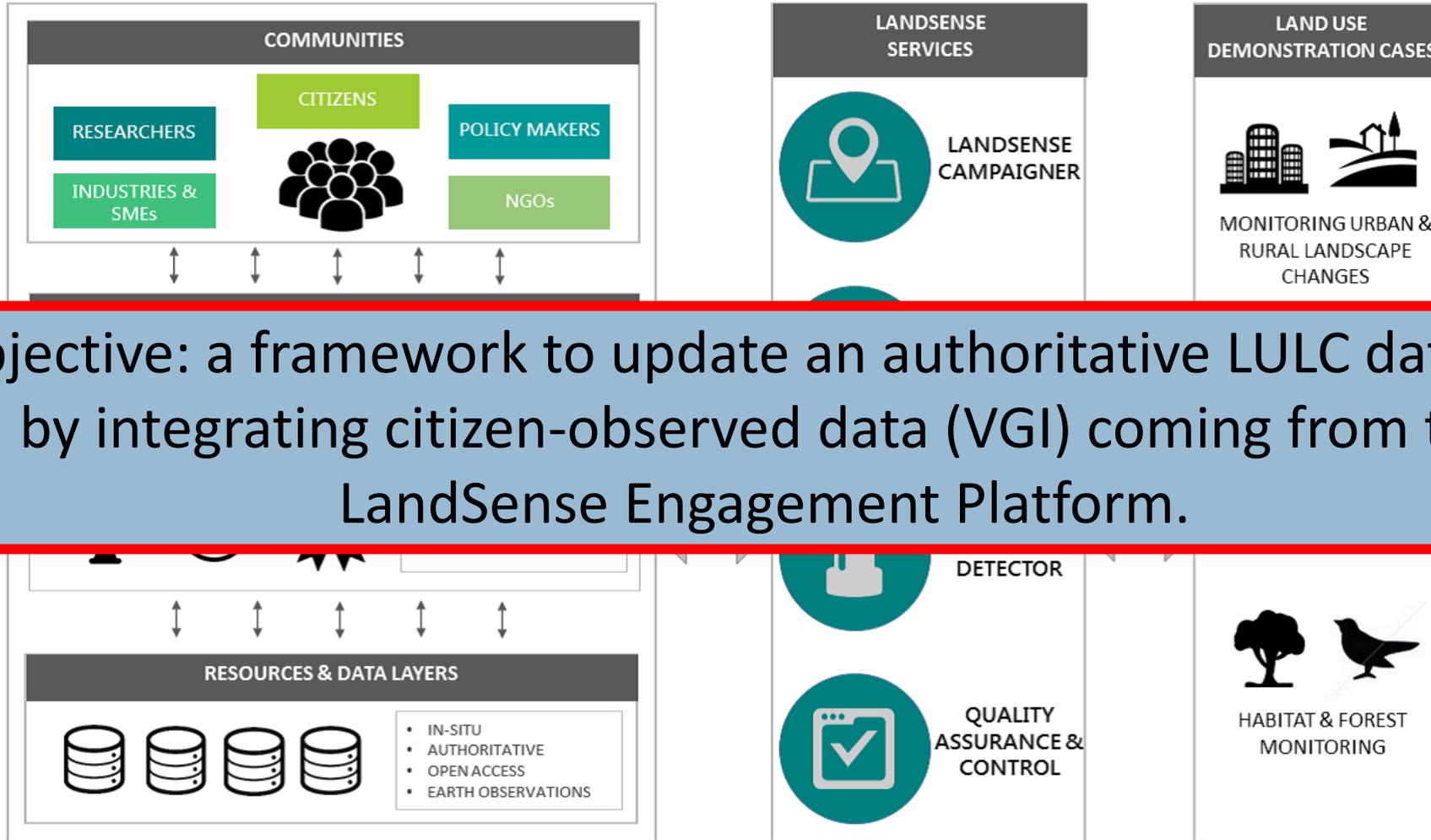
- The increasing demand for LULC data at fine spatial and temporal resolutions cannot be met by authoritative agencies such as a NMA



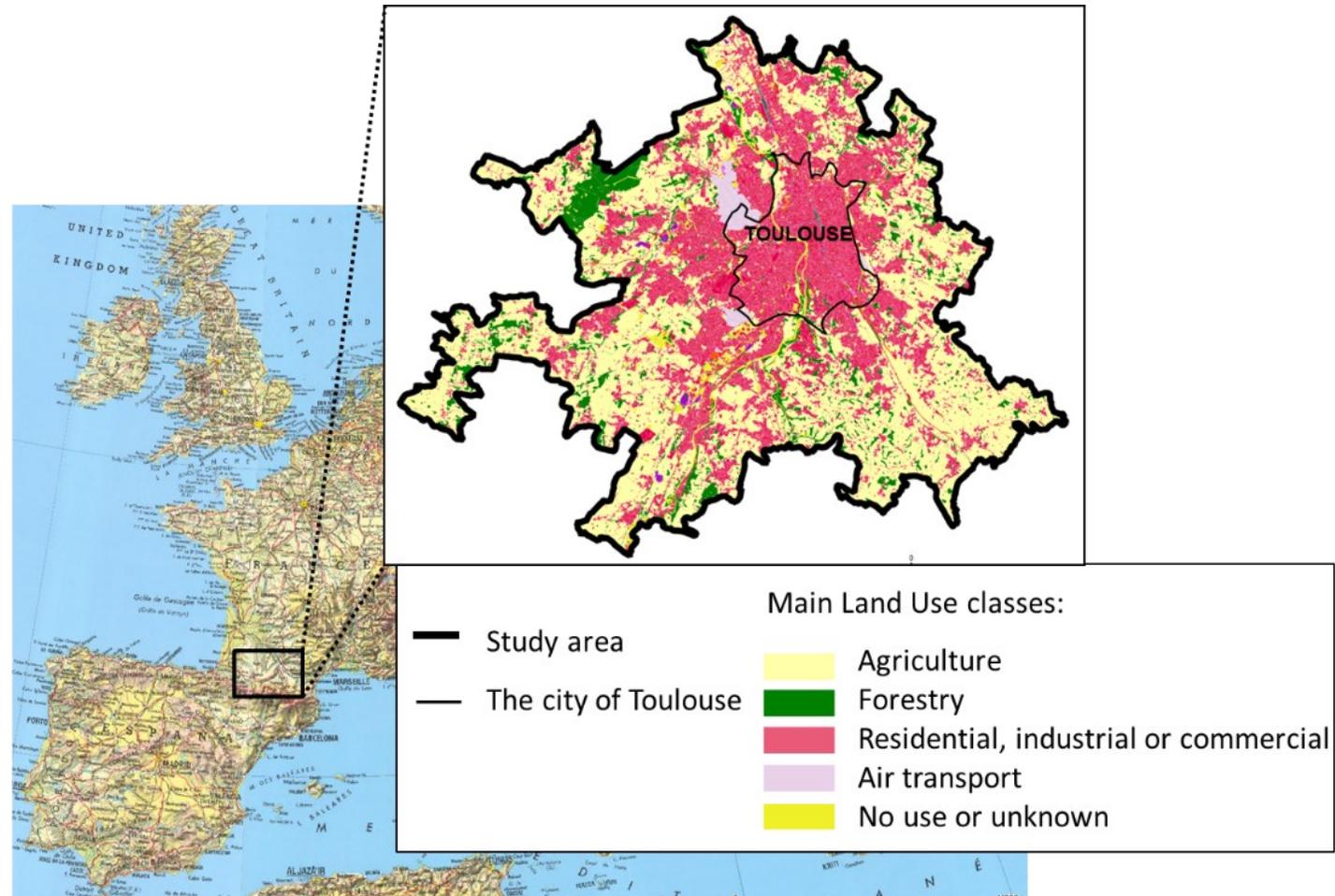
- Using citizen-sensors as a source of information



H2020 LandSense Project



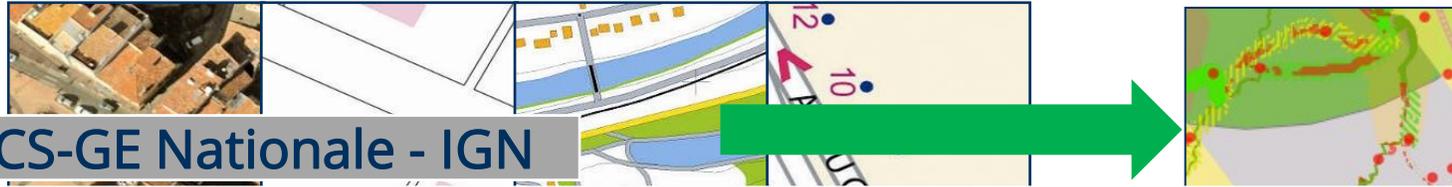
Study Area



Sources: Géoportail, GEOFLA

Toulouse pilot area in the south of France

OCS-GE



OCS-GE Nationale - IGN

Land Cover

CS1 Without vegetation		CS2 With vegetation	
CS1.1 Anthropized surface	CS1.2 Natural surface	CS2.1 Woody vegetation	CS2.2 Non-woody vegetation
...

Orthophoto Cadastral Topographic Address LULC (2016)

Land Use

US1 Primary production	US2 Secondary production	US3 Tertiary production	US4 Logistics transport networks and infrastructures	US5 Residence	US6 Other usage
...



■ US6.1
■ US6.3



■ CS2.2.1

- Authoritative LULC data (2016) produced by IGN-France.
- Vector data, non-overlapping polygons describing LULC.
- The nomenclature is INSPIRE-compliant, contains 17 LU classes and 14 LC classes.
- Updated on a three-year cycle.

Multi-source VGI - Tools

- In-situ* campaigns: **PAYSAGES mobile app** (available in the Apple Appstore and the Google Playstore)
- Web campaigns: **Paysages web application** (<https://paysages.ign.fr/>)

PAYSAGES web

PORTAL LANDSENSE

MON COMPTE

PAYSAGES

PARTICIPEZ AU SUIVI DE L'USAGE ET DE LA COUVERTURE DU SOL

ACCUEIL DONNÉES SIGNALEMENTS SUJETS CONFIGURATION

La maîtrise de la consommation des espaces naturels et agricoles est devenue un enjeu majeur pour notre société. Cette lutte contre l'étalement urbain ne peut se faire sans un suivi fin et régulier de l'occupation du sol sur les territoires concernés. Compte-tenu des enjeux qu'il représente, ce suivi est l'affaire de tous.

Devenez les sentinelles des paysages !
La plateforme collaborative Paysages propose à chacun de prendre une part active dans ce suivi.

EXPLORER LES DONNÉES

CONTRIBUER

CONFIGURER MON COMPTE

Accédez à la structure et au contenu des bases de données bâtiment et

Participez à l'enrichissement et à la validation des données d'occupation du

Gérez vos informations personnelles, vos groupes, vos profils...

PAYSAGES mobile

Paysages

Devenez les sentinelles des paysages!

SUIVANT

PAYSAGES wiki

PAYSAGES Wiki

Accueil Discussion Lire Modifier

Accueil

Sommaire

Tableau de bord

Navigation

Modifications récentes

Outils

Pages liées

Suite des pages liées

Pages spéciales

Version imprimable

Lien permanent

Information sur la page

Chercher les propriétés

Accueil

Le projet LandSense a comme objectif d'améliorer la connaissance et le suivi de l'information environnementale en proposant une plateforme à chercheurs, industriels et association non gouvernementale d'avoir accès aux données de l'usage et la couverture des sols, aux outils permettant différents services

- IGN producteur de données
- Évaluer la pertinence de la contribution citoyenne et communautaire sur la thématique OCS-GE

Pourquoi contribuer? [modifier]

La réduction de l'artificialisation des sols est un levier essentiel dans la protection de la biodiversité et la lutte contre le changement climatique: le ministre de la transition écologique et solidaire en a fait un de ses cheval de bataille. La France dispose d'objectifs climatiques ambitieux et ne peut sans utiliser les capacités naturelles des sols et, notamment, des terres agricoles.

Ce suivi est l'affaire de tous, devenez les sentinelles (veriteurs) des paysages et de la nature!

Pourquoi faire? [modifier]

- Enrichir le bâti, Description du bâti
- Enrichir et mesurer l'OCS-GE, Description d'OCS-GE

Où? [modifier]

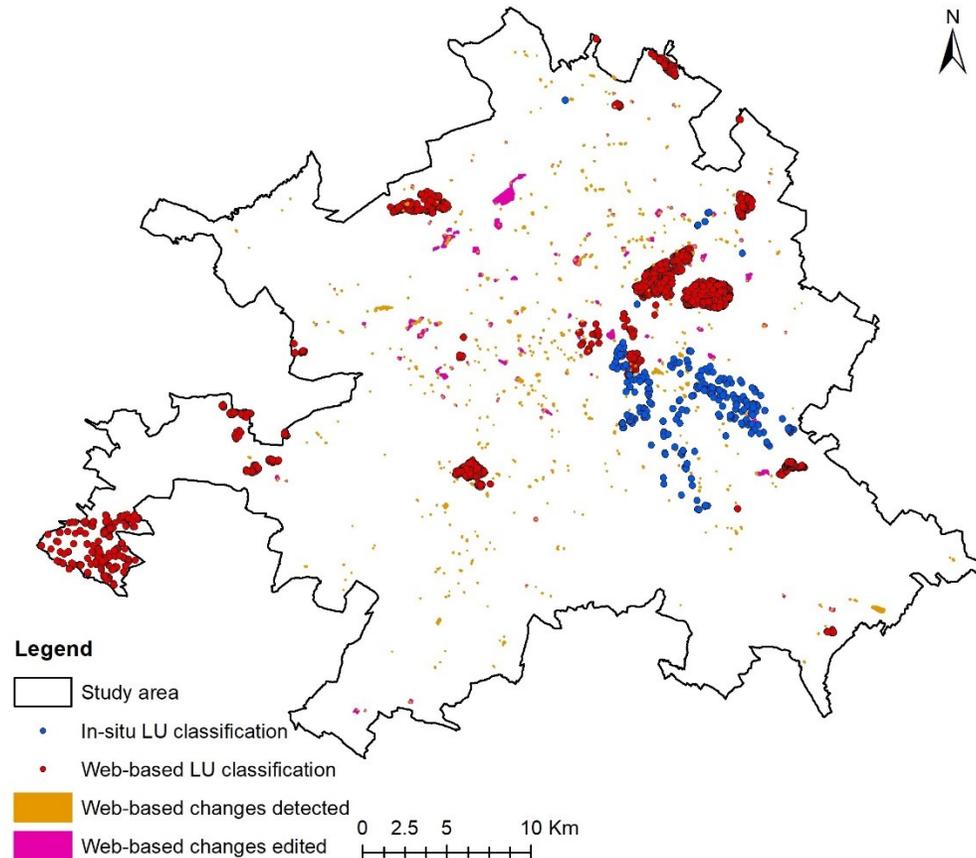
A ce niveau de notre challenge, nous focalisons notre campagne sur trois zones de la région d'Occitanie.

- Parc naturel Régional des Grandes Causses
- Agglomération toulousaine
- Pays Sud Toulousain

Comment puis-je aider? [modifier]

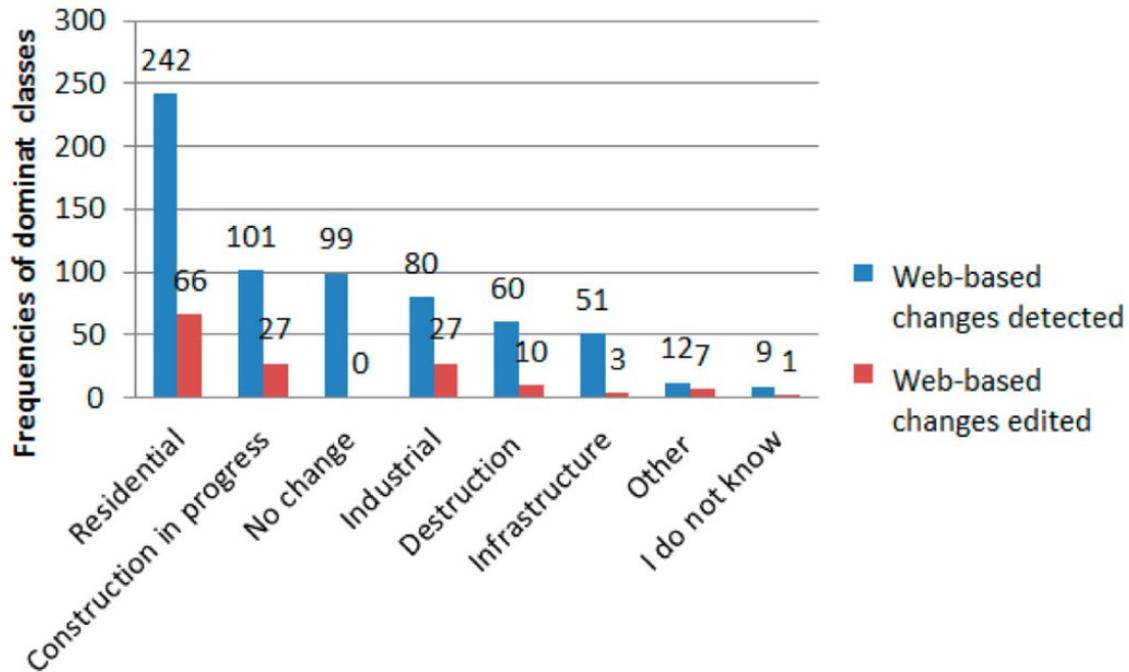
Vous êtes utilisateurs de la donnée, passionnés de la nature, membres de différentes communautés de protection de la nature, étudiants... vous

Multi-source VGI



- Multiple observations of LULC change detected via the web-based application (**Web-based changes detected**)
- Single observations of change edited by the users via the web-based application (**Web-based changes edited**)
- Single observations of LU classification obtained from the web-based application (**Web-based LU classification**)
- Multiple observations of LU classifications obtained from the mobile application (**In-situ LU classification**)

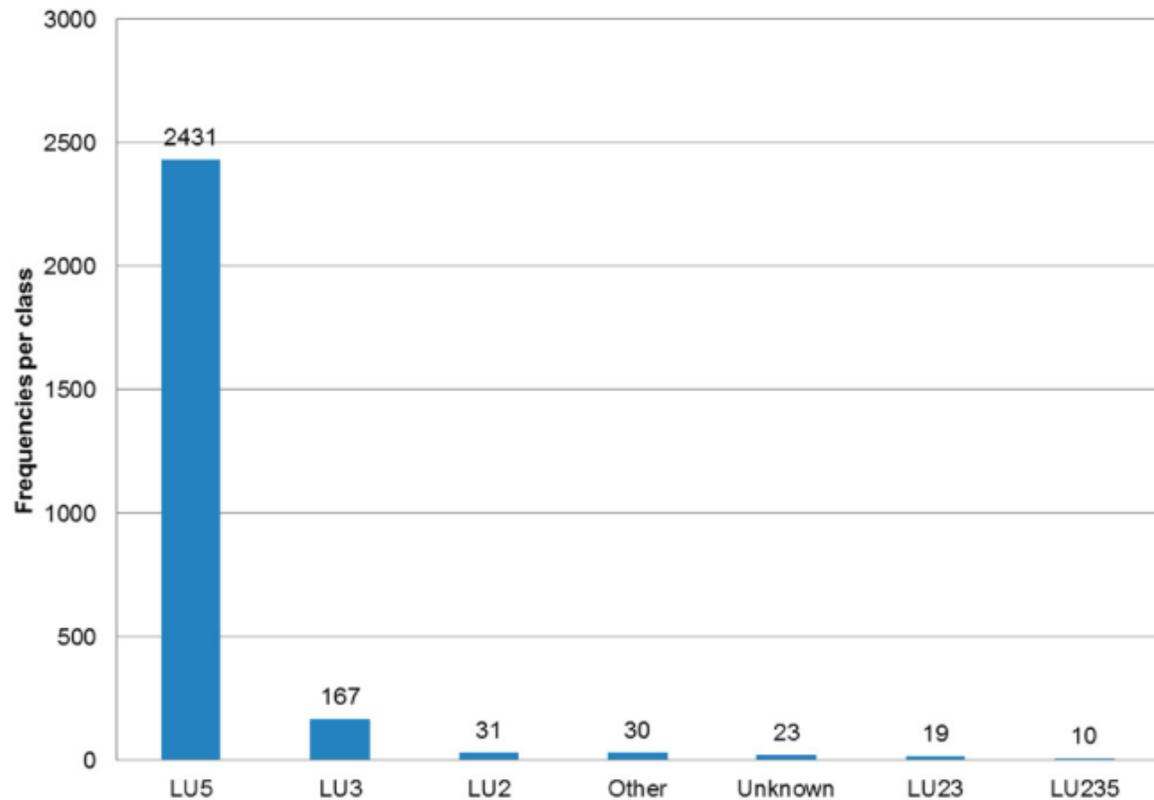
Multi-source VGI



Classification of change detection observations

- Multiple observations of LULC change detected via the web-based application (**Web-based changes detected**)
- Single observations of change edited by the users via the web-based application (**Web-based changes edited**)
- Single observations of LU classification obtained from the web-based application (**Web-based LU classification**)
- Multiple observations of LU classifications obtained from the mobile application (**In-situ LU classification**)

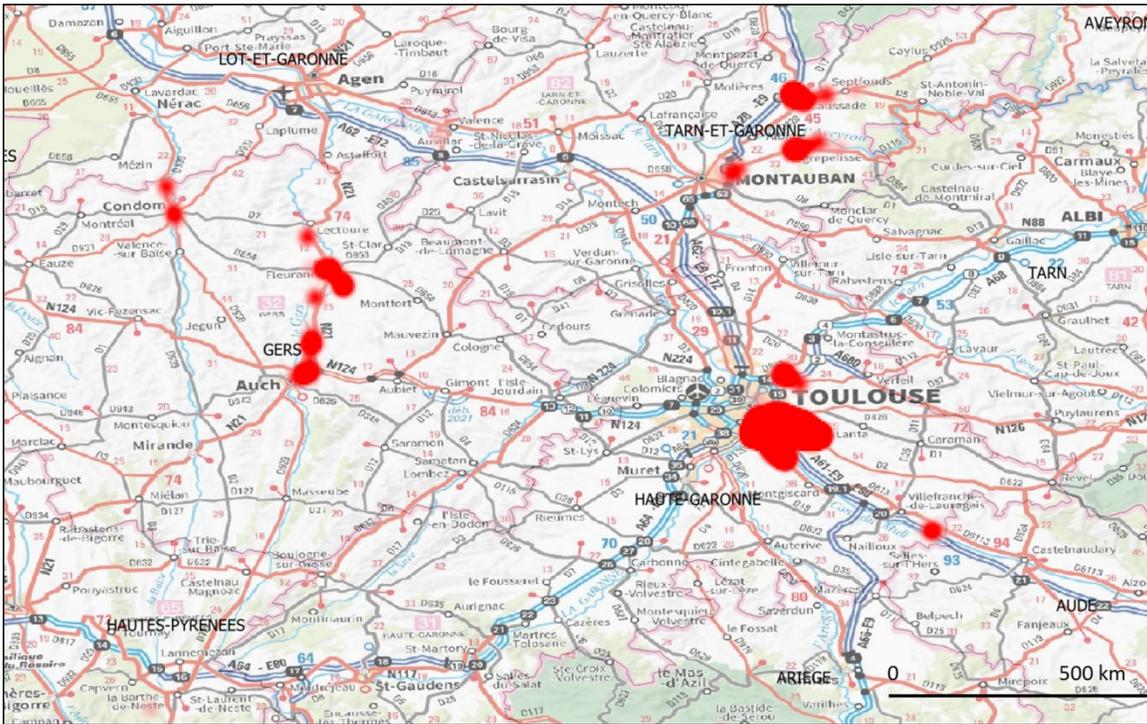
Multi-source VGI



Distribution of LU classification chosen by contributors to distinguish the combined classification LU235 (*Residential, Commercial and Industrial*)

- Multiple observations of LULC change detected via the web-based application (**Web-based changes detected**)
- Single observations of change edited by the users via the web-based application (**Web-based changes edited**)
- Single observations of LU classification obtained from the web-based application (**Web-based LU classification**)
- Multiple observations of LU classifications obtained from the mobile application (**In-situ LU classification**)

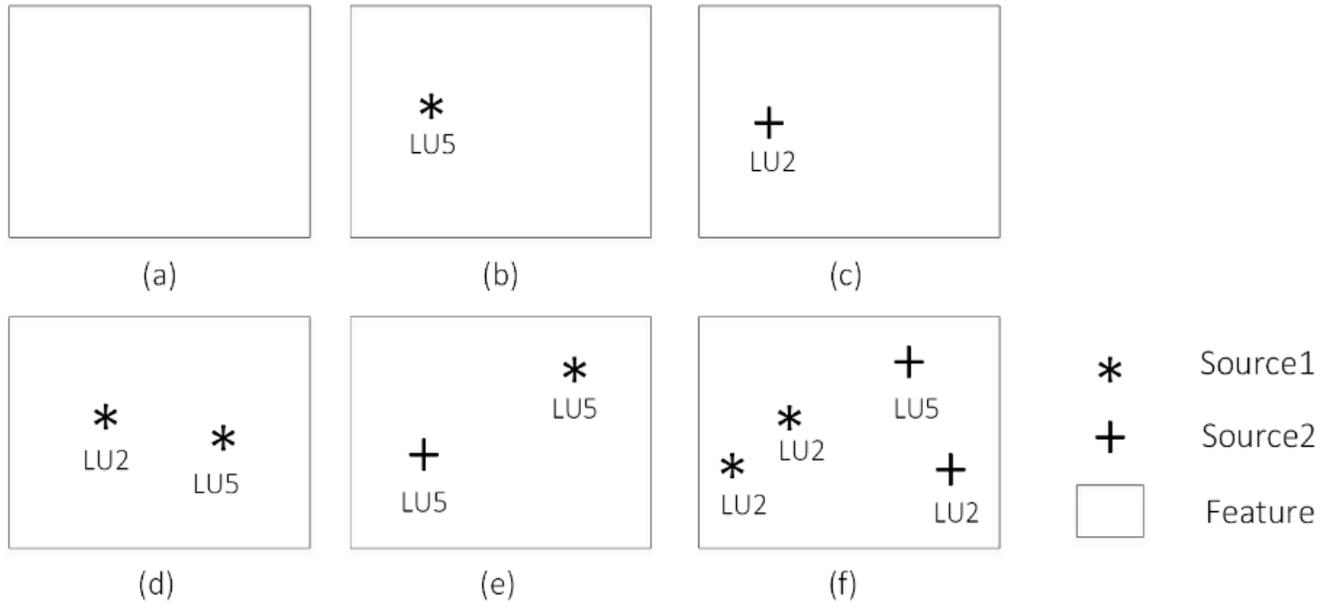
Multi-source VGI



The spatial density of locations using heatmap visualization
304 locations were visited in the field and 369 observations were collected.

- Multiple observations of LULC change detected via the web-based application (**Web-based changes detected**)
- Single observations of change edited by the users via the web-based application (**Web-based changes edited**)
- Single observations of LU classification obtained from the web-based application (**Web-based LU classification**)
- Multiple observations of LU classifications obtained from the mobile application (**In-situ LU classification**)

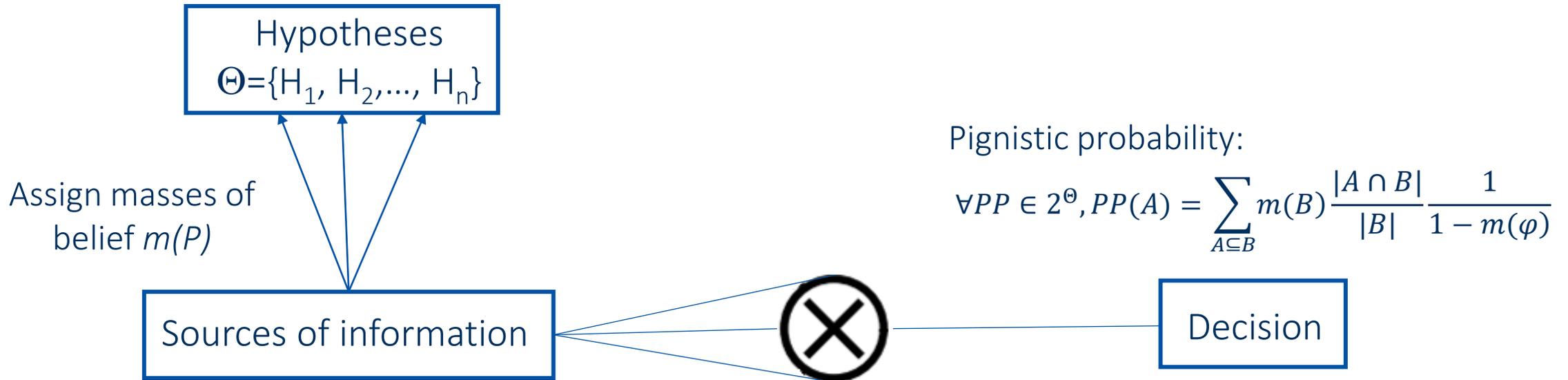
Issue



- Each feature (OCS-GE polygon) is characterised by zero, one or more observations coming from different data sources
- Citizen-observed data correspond to different types of geometry and have different values of change information
- How to merge them -> Dempster-Shafer Theory (DST)

■ Dempster-Shafer Theory (DST)

DST offers tools to combine several sources of information such as the **Dempster's rule** of combination. It can manage different types of values to make a final decision, generating confidence related to each value.



Conjunctive Dempster's rule:

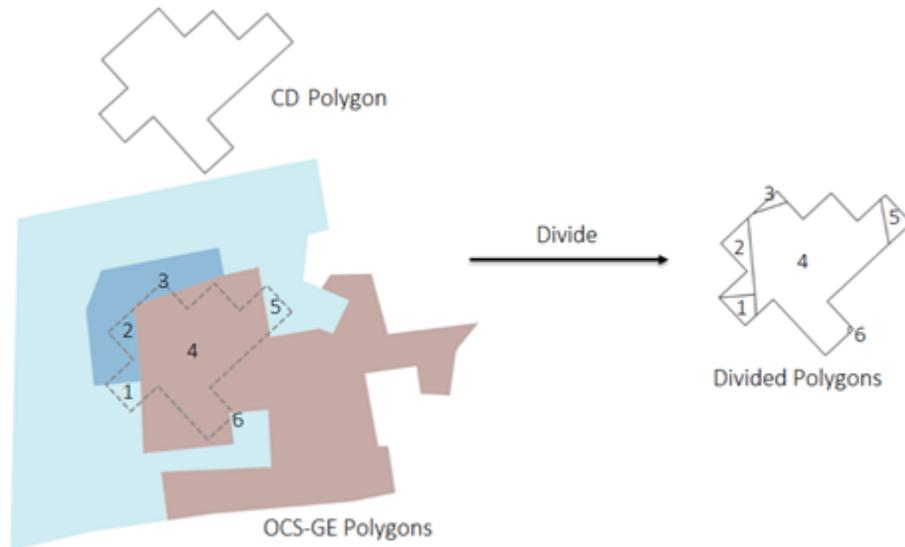
$$\forall P \in 2^\Theta, m_{12}(P) = \frac{\sum_{P' \cap P'' = P} m_1(P') * m_2(P'')}{1 - \sum_{P' \cap P'' = \varphi} m_1(P') * m_2(P'')}, P' \text{ and } P'' \in 2^\Theta$$

Workflow

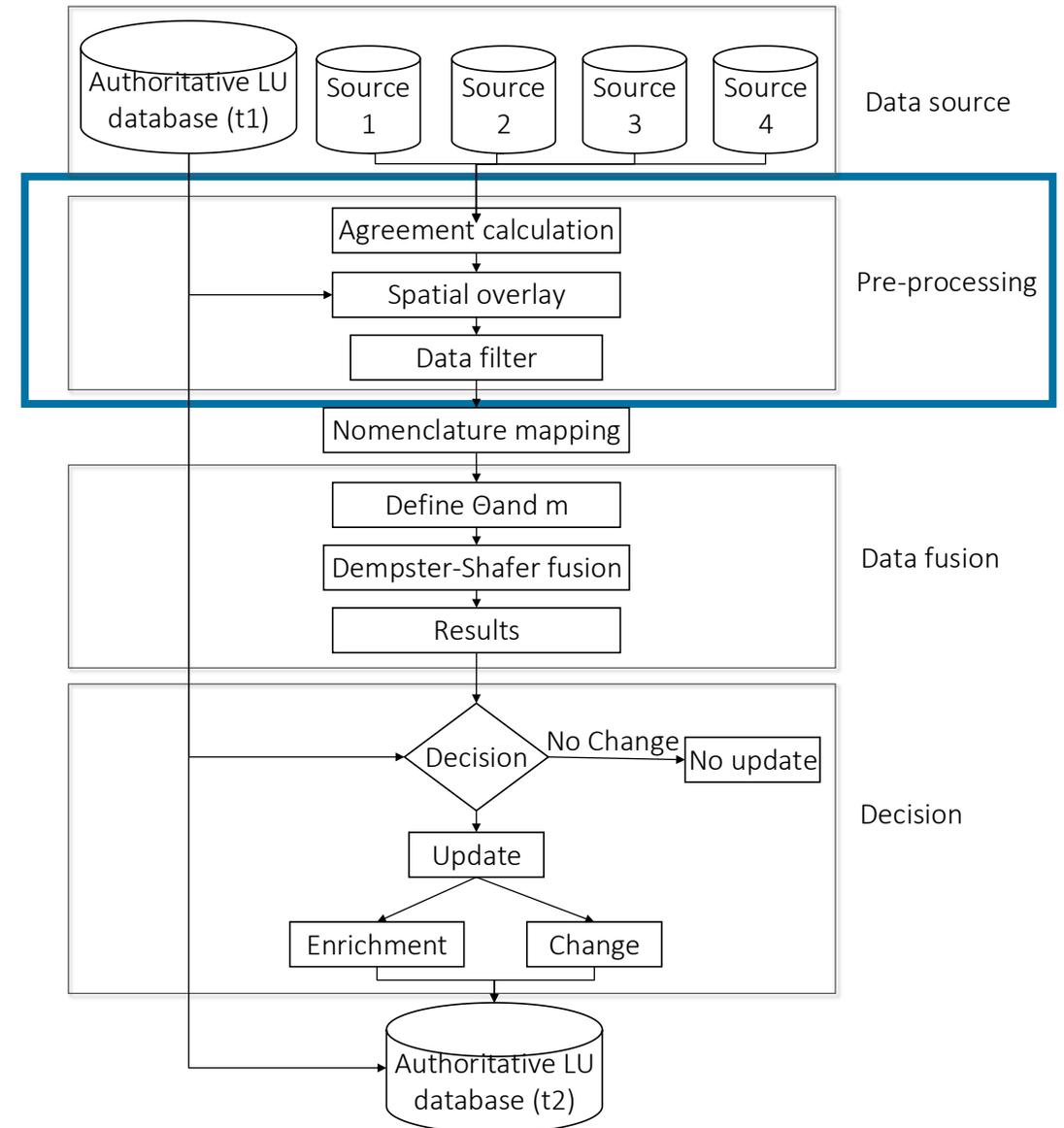
Data Preprocessing

Agreement calculation

Spatial overlay



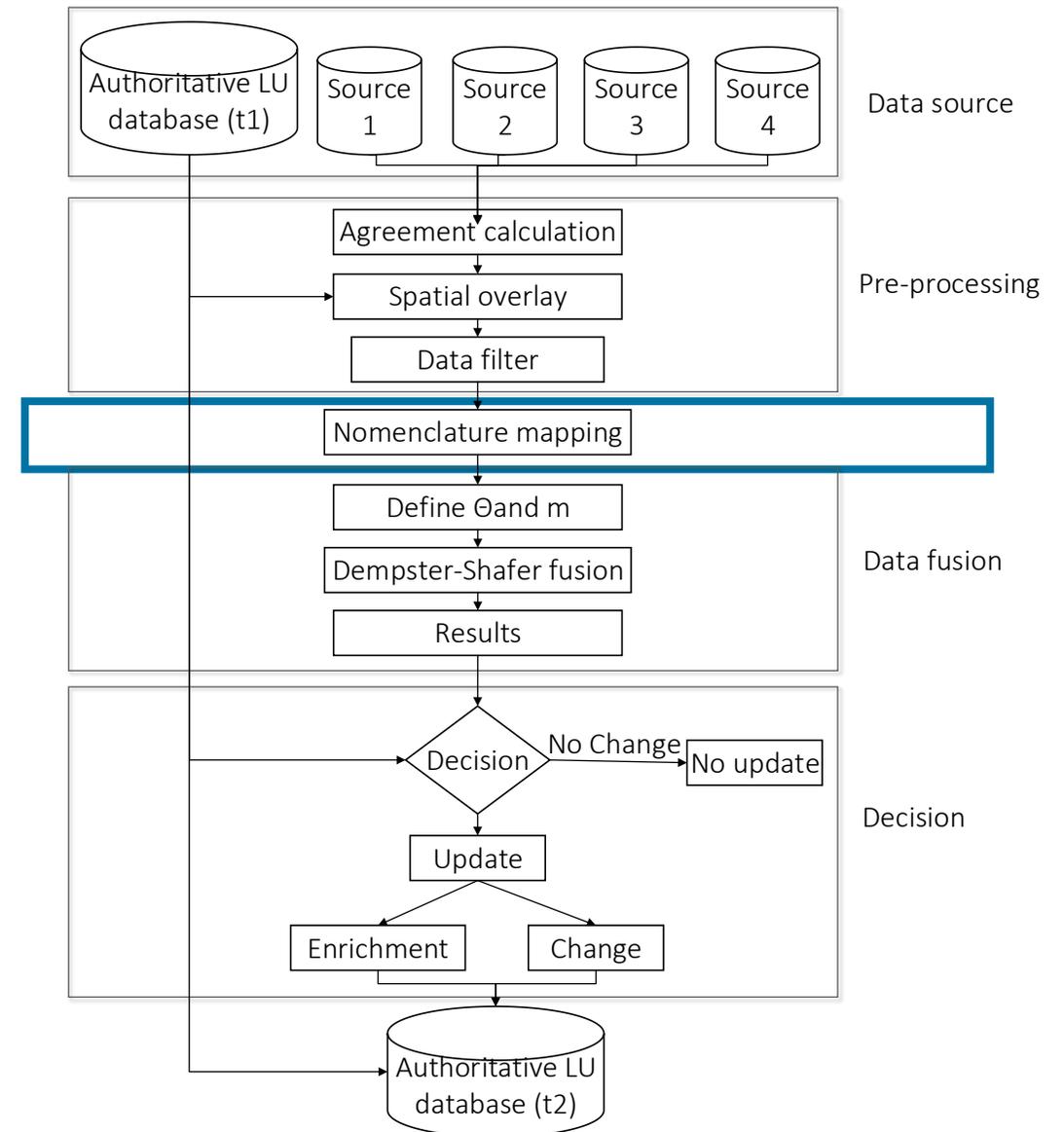
Data Filter



Workflow

- Data Preprocessing
- Mapping LU labels to OCS-GE nomenclature

Label for change features	OCS-GE LU nomenclature
<i>Residential</i>	LU5
<i>Industrial</i>	LU2
<i>Infrastructure</i>	LU3
<i>Other</i>	LU6
<i>Destruction</i>	LU6
<i>Construction in progress</i>	LU6.1
<i>No change</i>	None
<i>I do not know (skip)</i>	None



Workflow

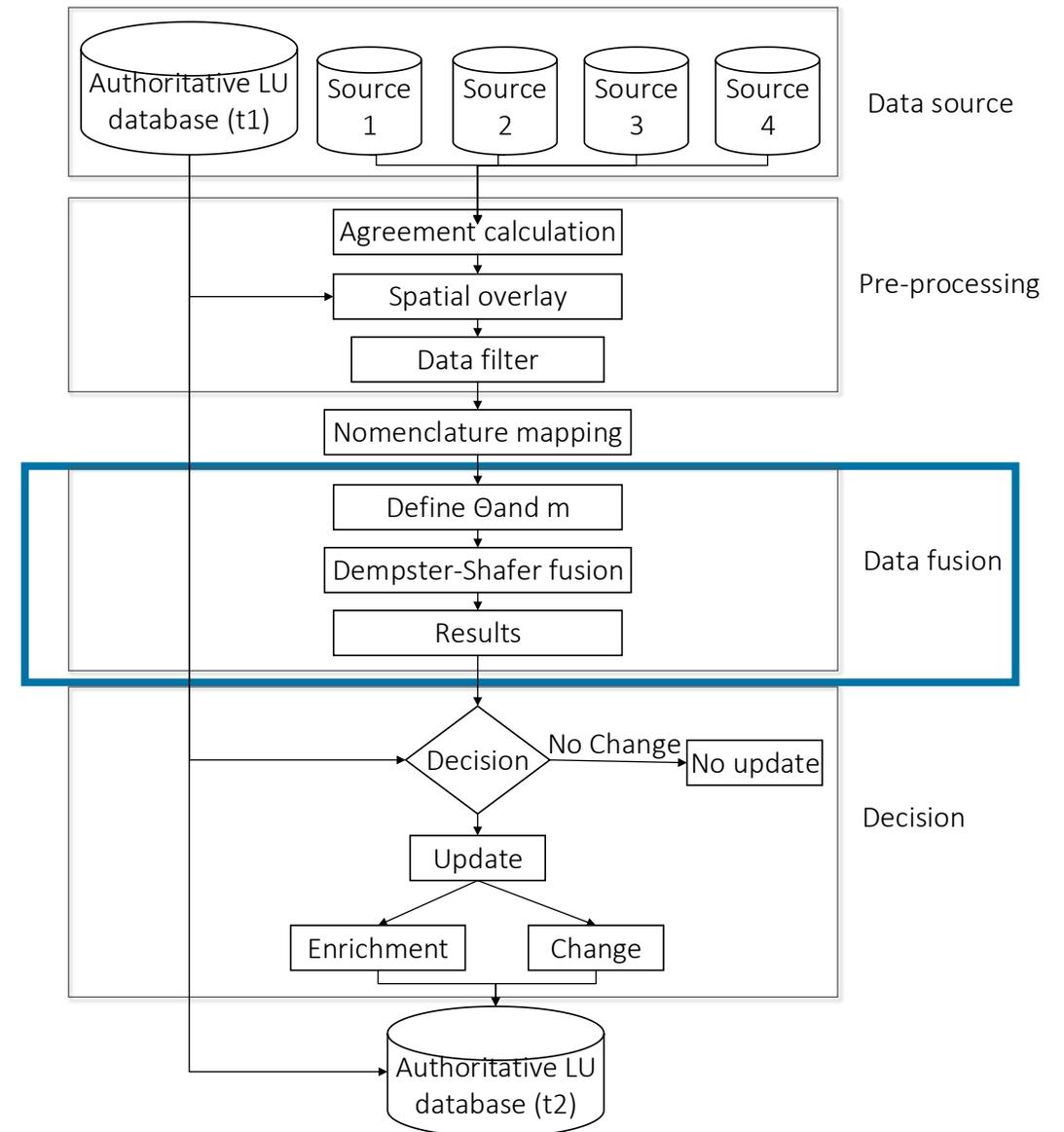
- Data Preprocessing
- Mapping LU labels to OCS-GE nomenclature
- Data fusion via Dempster-Shafer Theory

- Define Θ and m

$$\Theta = \{LU1.1, LU1.2, \dots, LU2, LU3, LU4.1.1, \dots, LU5, LU6.1, \dots, LU6.3\}$$

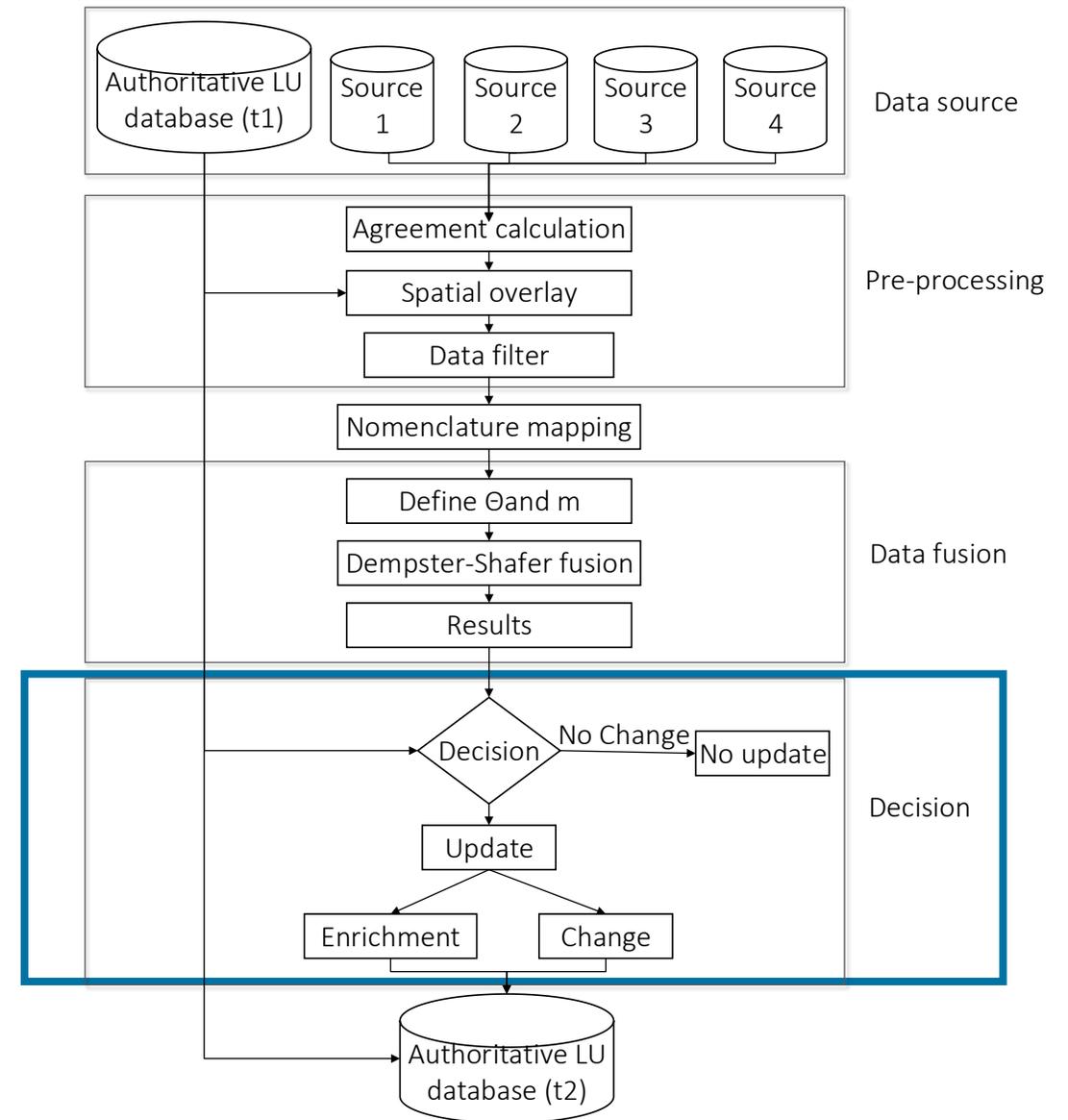
$$m(P) = Agreement * \frac{Area(CDS)}{Area(OCS-GE)}$$

- Data sources combination
- DST-derived results: the criteria of maximum of pignistic probability

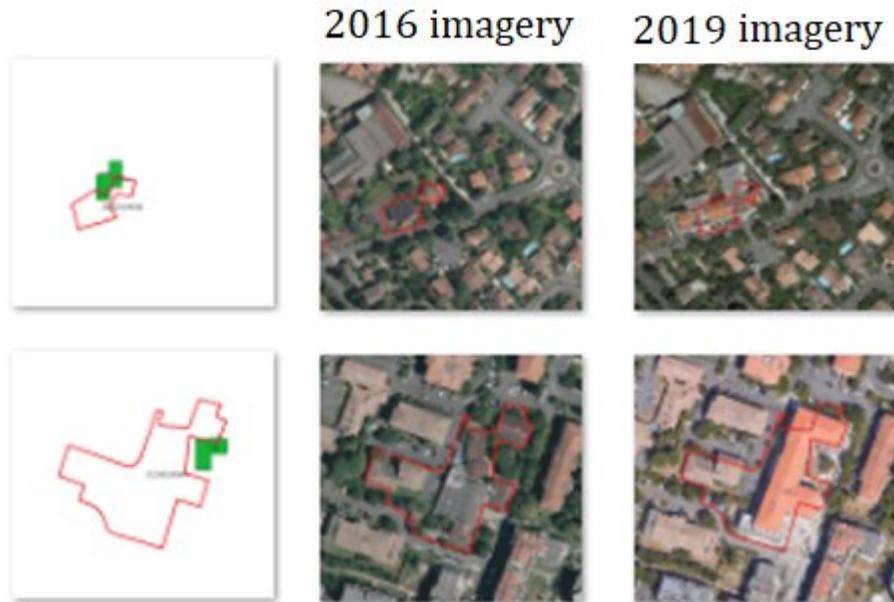


Workflow

- Data Preprocessing
- Mapping LU labels to OCS-GE nomenclature
- Data fusion via Dempster-Shafer Theory
- Decision step



Manual validation

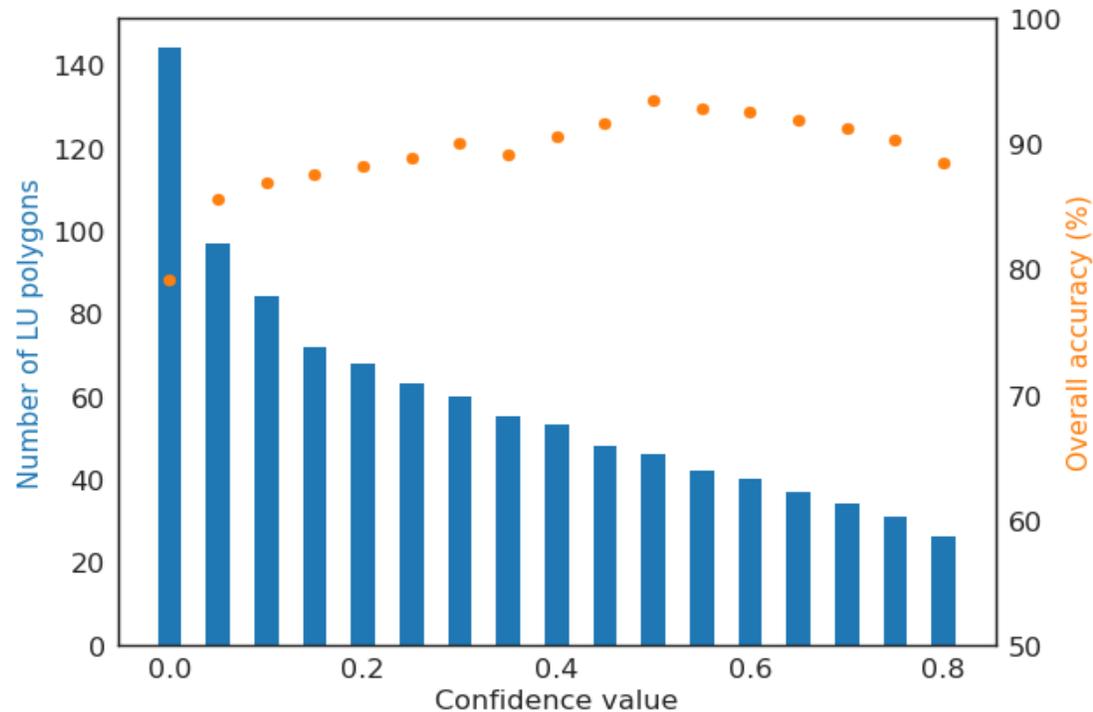


Imagery from 2016 and 2019 for each feature



Google street view imagery

DST results

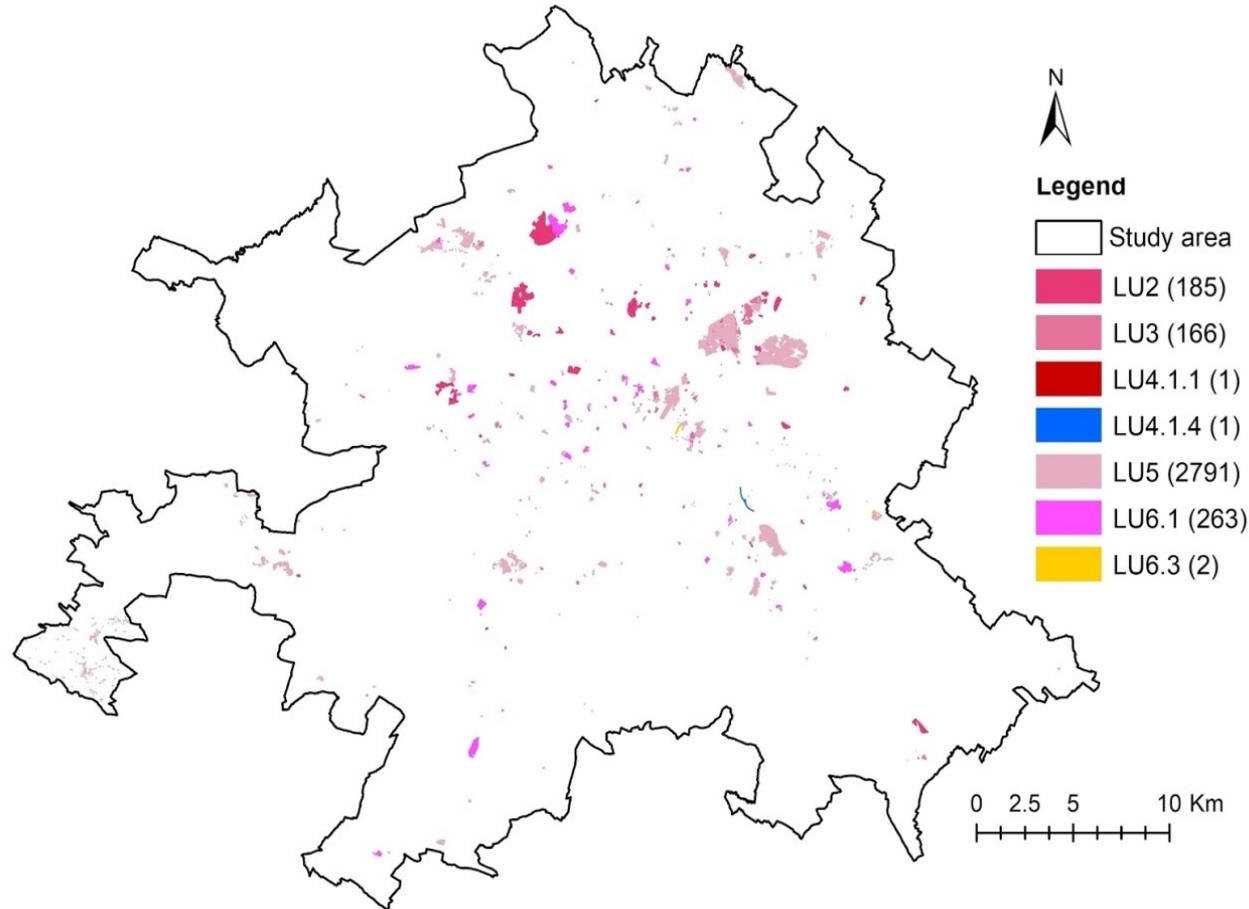


The overall accuracy and the number of LU polygons considered by confidence threshold values for 144 features having two or more contributions

		Ground truth				Total
		LU2	LU5	LU6.1	LU6.3	
DST-derived solution	LU2	17	1	1	0	19
	LU5	0	47	3	1	51
	LU6.1	0	7	19	0	26
	LU6.3	0	1	0	0	1
	Total	17	56	23	1	

Confusion matrix for features having two or more contributions with a confidence threshold greater than 0.05 (OA=85.6%)

■ Updating the authoritative land use database

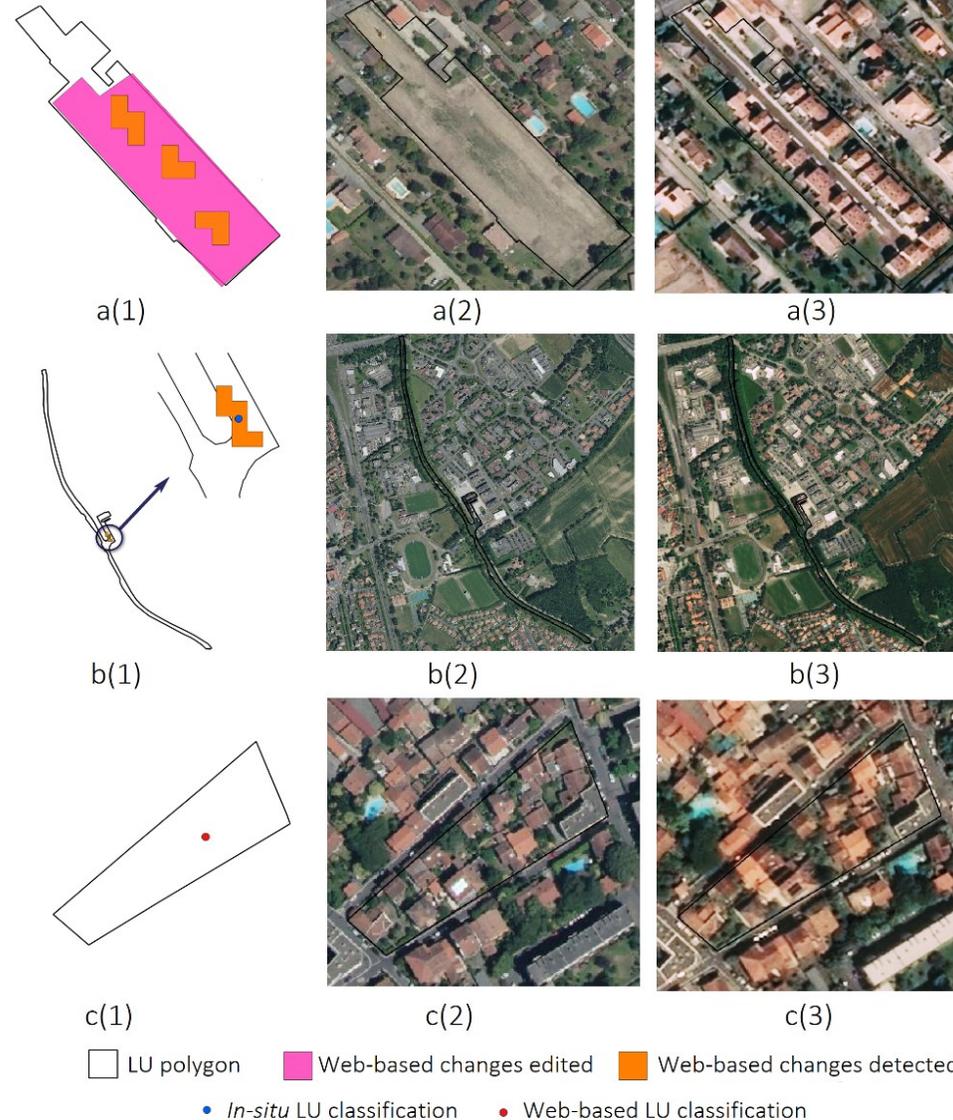


The LU polygons updated with the corresponding new LU labels. (LU2-Industrial, LU3-Commercial, LU4.1.1-Road transport, LU4.1.4-Water transport, LU5-Residential, LU6.1-Construction zones, LU6.3-Not currently used)

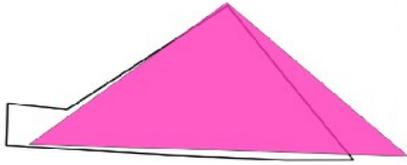
- 3409 LU polygons under consideration.
- The decision is classified into three cases: no update, enrichment, and change.
- 25 LU polygons (0.1 km²) were classified as no-update.
- 409 LU polygons (7.36 km²) were classified as change.
- 2975 LU polygons (5.75 km²) as enrichment.

Examples

- (a) change from LU6.3 (*Not currently used*) to LU5 (*Residential*)
- (b) no update (LU4.1.4 - *Water transport*)
- (c) LU235 (*Industrial, Commercial and Services, and Residential*) to LU5 (*Residential*)



Error Analysis



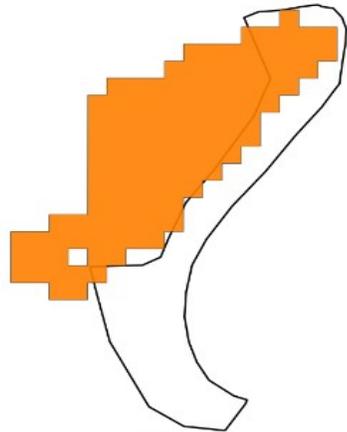
a(1)



a(2)



a(3)



b(1)



b(2)



b(3)

LU polygon
 Web-based changes edited
 Web-based changes detected

- Knowledge gap of the volunteers regarding the authoritative land use nomenclatures
- Partial change caused by inaccurate or modified polygon geometry

■ Advantages

The proposed workflow integrates authoritative LULC data with heterogeneous VGI based on Dempster-Shafer theory.

- Taking into account information coming from different and heterogeneous data.
- The scalability of the framework.

■ Disadvantages

- The spatial distribution of VGI.
- The quality of the VGI would also need to be taken into account in the framework.
- The geometry of the LULC polygons is not modified.

■ Conclusion

- The proposed framework integrates authoritative LU data with heterogeneous multi-source VGI based on Dempster-Shafer theory.
- Multi-source VGI to update the authoritative LU database with an overall accuracy of 85.6% .
- By taking advantage of the scalability of the proposed framework, potential or existing data sources could be designed or directly added as inputs to the fusion process.

Thanks for your attention !
