



<https://www.didip.eu>
@DiDip_ERC@fedihum.org

UNIVERSITY OF GRAZ

Department of
Digital Humanities



From maps to Document: Controlling Urbanity in Large Charter Corpora

Workshop - Digital Methods in Diplomatic Research

Giuseppe Consolo - University of Graz - 13/02/2026

Monasterium.net corpus

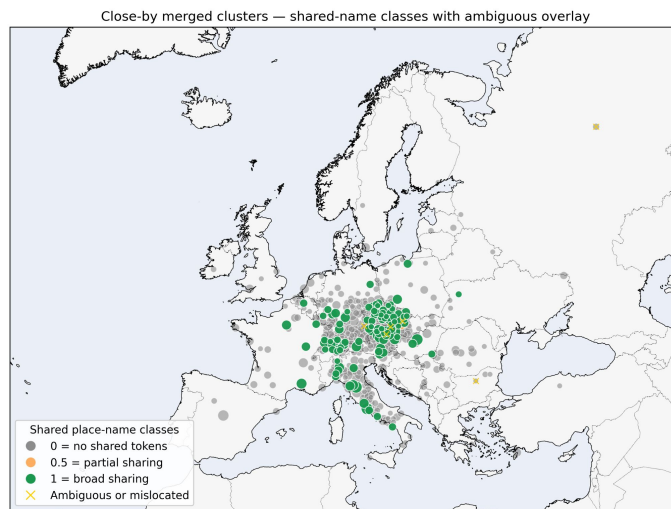
Extraction
place_of_issuing



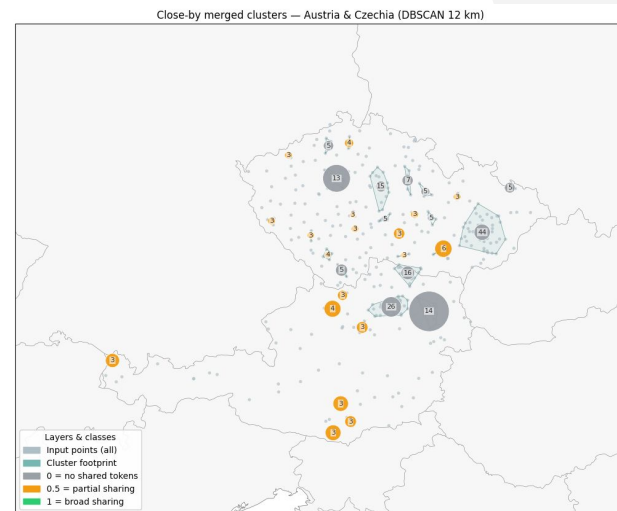
Automatic Match with
Geonames



Creating Maps



Issuing places: close-by clusters(DBSCAN)



Issuing places in Austria/Czechia merged by proximity(DBSCAN 12 km); circle=total occurrences, label=points(n), colour=shared place-name class


From place of issuing to “urban”: a fragile deduction

Datatio topica

*Diser brief ist geben
ze Wienn...*



Coordinates

Wien,
48.20849, 16.37208 



Urban Level?

Uncontrolled
step

From place of issuing to “urban”: a fragile deduction(2)



url	language	PlaceName
https://www.monasterium.net/mom/AT-9:AG/Goettweig058/1500_IX_27/charter	german	Mautern
https://www.monasterium.net/mom/AT-9:AG/Goettweig058/1456_VII_15/c charter	german	Mautern in Steiermark



PlaceName	in_atlas_1km	in_atlas_5km
Mautern	FALSE	TRUE
Mautern in Steiermark	FALSE	FALSE



Urbanity Level
Urban
Urban
Urban context
ambiguous
non-urban

normalization

coverage

choice

How can we control urbanity, in a large charter corpus, instead of assuming it?

1

Document identification
- atomid
- id
- language
- date
- placeName
- modernPlaceName

2

Similarity & clustering (FLAME)
- flame_cluster_id
- cluster_size
- is_singleton

3

External spatial signals
- in_atlas_1km / 5km
- matched_city_1km / 5km

4

Diplomatic interpretation (manual)
- Indicators
- Confounder
- Urbanity Level
- Note (rationale)



intersection
table

Document identification

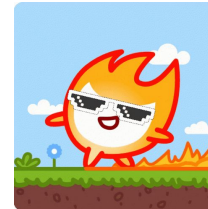
1

Document identification

- atomid
- id
- language
- date
- placeName
- modernPlaceName

atomid	id	language	date	placeName	modernPlaceName
tag:www.r	AT-WStLA	german	XV	Eisenstadt	Eisenstadt
tag:www.r	AT-OOeLA	german	XIV	Enns (Ens)	Pyburg
tag:www.r	AT-StiAV_	german	XV	Graz	Graz
tag:www.r	AT-StiAK_	german	XIV	Klosteneuburg	Klosterneuburg

How and what clustering gives us FLAME



Kovács, T. (2025). *FLAME: Formulaic Language Analysis in Medieval Expressions* (Version 1.0.0) [Computer software]. GitHub.
<https://github.com/kreedit/FLAME>

FLAME HTML output (excerpt)

Comparison: AT-WstLA_HAUrk_356_tenor.txt → AT-WstLA_HAUrk_355_tenor.txt

Cosine Similarity: 0.7401

File 1: AT-WstLA_HAUrk_356_tenor.txt

Ich Nicla der Paltram zu den zeiten perigmaister der Hochgeporn Fuerstin Herzogin Johanna in Oesterreich in Steyr und in Kernten vergich offenleich an disem prief daz fuer mich cham in die schrann ze medlinch do ich saz an rechtem gericht her Haunolt der Schuechler purger ze wienn und chlagt mit vorsprechen hincz **Leutolts** weingarten des **Juden sun von perichtolzdorf**: der do leit in dem prunner pach umb zwai phunt versessens purchrecht und umb alle die zwispil di mit recht dar auf gegangen sind als sein prief sagt als verre daz ich im poten dar auf gab Hainreichen den werder Merten den Sulczer Eberharten den Choelblein di habent dar umb gesagt vor rechtem gericht mit iern trewn als si ze recht schollten: daz seu den weingarten geschaut und geschaczt haben daz er so teur nicht enist als daz versessen purchrecht und di zwispil di mit recht dar auf ertailt sind und ward gefragt waz recht wer do geviel im do mit frag und mit uertail er scholt mier mein recht geben daz hat er getan und scholt ich in des vorgeantent weingarten **gevaltig** machen und an di guber gewer setzen daz han ich auch getan fuerbas ledichleich und vreichlich ze haben und allen seinen frum da mit ze schaffen verchaffuen **versetzen** und **gewen** swem er welle an allen

File 2: AT-WstLA_HAUrk_355_tenor.txt

Ich Nicla der Paltram zu den zeiten perigmaister der Hochgeporn Fuerstin Herzogin Johanna in Oesterreich in Steyr und in Kernten vergich offenleich an disem prief daz fuer mich cham in di schran ze Medlinch, do ich saz an rechtem gericht her Haunolt der Schuechler purger ze wienn und chlagt mit vorsprechen hincz **Niclas** weingarten des **Emrinch** der do leit in der **Pachlukchen ze Prunn** umb ain phunt versessens purchrecht und umb alle di zwispil di mit recht darauf gegangen sind als sein prief sagt als verre daz ich im poten darauf gab Hainreichen den Werder Merten den Sulczer Eberharten den Choelblein di habent darumb gesagt vor rechtem gericht mit iern trewn als si ze recht scholten daz seu den **vorgenantent** weingarten geschaut und geschaczt haben daz er so teur nicht enist als daz versessen purchrecht und di zwispil di mit recht dar auf ertailt sind und ward gefragt waz recht wer do geviel im do mit frag und mit uertail er scholt mier mein recht geben daz hat er getan und scholt ich in des vorgeantent weingarten **gewaltig** machen und an die guber gewer **setzen** daz han ich auch getan fuerbaz ledichleich und vreichlich ze haben und allen seinen frum damit ze schaffen verchaffuen **versetzen** und **geben** swem er welle an allen iersal . Und des

Similarity signal (cosine) is explicit and inspectable. It is not an urbanity label
Clusters ≠ cities
(documents can share formulae across contexts)

Similarity and Clustering

2

Similarity & clustering (FLAME)
- flame_cluster_id
- cluster_size
- is_singleton

id	flame cluster id	cluster size	is singleton
AT-WSLA HAUrk 2978 tenor.txt	SINGLETON	1	True
AT-OOeLA GleinkOSB 1337 IV 24 tenor.txt	SINGLETON	1	True
AT-StaAZ Urkunden 45 tenor.txt	C004	31	False
AT-WSLA HAUrk 154 tenor.txt	C042	5	False
AT-WSLA HAUrk 2905 tenor.txt	C002	53	False

External spatial signals

3

External spatial signals
- in_atlas_1km / 5km
- matched_city_1km / 5km

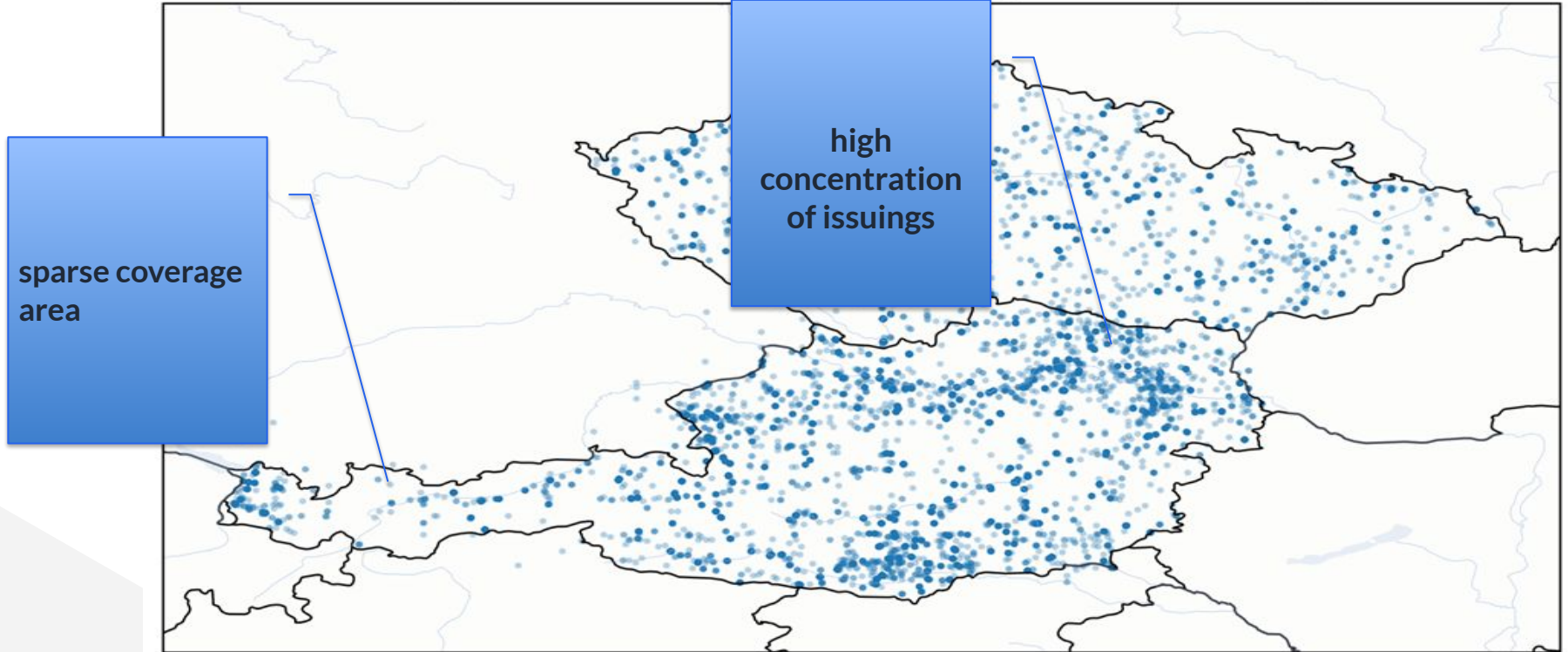
ÖSTERREICHISCHER STÄDTEATLAS



in_atlas_1km	in_atlas_5km
FALSE	TRUE
FALSE	TRUE
FALSE	TRUE
TRUE	TRUE
TRUE	TRUE

External spatial signals

Map A1 — Baseline place-of-issuing (Austria + Czechia)



A controlled reference standard

External reference

atlas
(1 km / 5 km)
+
matched_city (1 km / 5 km)

Internal reference

seedset + diplomatic
interpretation
(Indicators / Confounder)
+
Urbanity Level (manual)

intersection

Seedset

		External reference (atlas)	
		match	no match
M a n u a l u r b a n i t y	Urban (manual)	<p>15</p> <p>A: urban + atlas</p>	<p>15</p> <p>B: urban - atlas</p>
	Not urban (manual)	<p>10</p>	<p>10</p>

id	modern	1km	5km	Urbanity	category
AT-WStLA_HAUrk_2978_tenor.txt	nan	False	True	Urban	A_TP_urban+atlas
AT-OOeLA_GleinkOSB_1337_IV_24_tenor.txt	Pyburg	False	True	Urban context	A_TP_urban+atlas
AT-StiAV_VorauCanReg_654_tenor.txt	Graz	False	True	Urban context	A_TP_urban+atlas
AT-StiAK_KlosterneuburgCanReg_1323_IX_29_tenor.txt	Klosterneuburg	True	True	Urban context	A_TP_urban+atlas
AT-StiASch_SchlierbachOCist_1322_V_25_tenor.txt	Klosterneuburg	True	True	Urban context	A_TP_urban+atlas

Crop: 5 rows from seedset file (xlsx)

C: atlas - urban

D: neither

Four levels of urbanity

Level	Definition (from seedset Note)	Cue (from seedset)
Urban	Urban certain because the charter is explicitly issued in the name of city authorities (civic role: issuer).	Burgermaister
Urban context	Urban context because the council and citizens of Enns are explicitly named as the legal forum/enforcement authority and the act is authe...	"Burger ze Ens"
ambiguous	Ambiguous because the issuer and a witness are labelled Bürger zu Stain (urban social label), but the charter is a private Grunddienst ar...	buerger/bürger zu Stain"
non-urban	Non-urban because the text is a notarial/ecclesiastical authentication of an imperial charter and contains no municipal institutions or o...	"burger zu Kalsching"

Urbanity Level is assigned manually; rationale is stored in Note + indicators/confounders.

Indicators & confounders

✓ Indicators (strong signals)

Micro-example (seedset): AT-WStLA_HAUrk_2978_tenor.txt

Burgermaister; Rat; Richter; Mitburger; offen
Rat

civic offices / council / citizens named explicitly
urban forum as acting authority (not only a place-name)
authentication linked to municipal institutions

⚠ Confounders (false cues)

Micro-example (seedset): AT-OOeLA_MondseeOSB_1291_X_06_tenor.txt

“ze Crems” (urban place-name used only as
dating/location, with no civic offices/institutions); the
entire charter is framed by ecclesiastical authorities
(bishop of Passau, abbot of Mondsee, parish structures)
and church rights.

datatio “in city X” without municipal action
seigneurial / ecclesiastical framing dominates
territorial rulers + city mention as logistics

First case: Expected match

Intersection row (highlighted)		Text fragment 1 (seedset Indicators)
id	AT-StaAZ_Urkunden_45_tenor.txt	“richter, rath unnd die burger gemeinlich unnserer statt Zwetl”
modernPlaceName	Vienna	
flame_cluster_id	C004	
cluster_size	31	Text fragment 2 (seedset Indicators / Note) “confirmation ... ir gnadt, freyhait, recht, privilegia ...”
in_atlas_1km	YES	
in_atlas_5km	YES	
Urbanity Level	Urban context	
category	A_TP_urban+atlas	

Second case:

FALSE POSITIVE

Intersection row (highlighted)	
id	AT-StiAHe_HerzogenburgCanReg_1323_III_12.1_tenor.txt
modernPlaceName	Vienna
flame_cluster_id	C012
cluster_size	13
in_atlas_1km	YES
in_atlas_5km	YES
Urbanity Level	non-urban
category	C_atlas_not_urban

Text fragment 1 (seedset Confunder)

Rural/feudal-ecclesiastical transaction (property to “gotshaus ze Nuzdorf”)

Text fragment 2 (seedset Confunder / Note)

enforcement clause “hintz Herzogenburch ... hintz eim erben wirt” refers to a lodging/host, not a civic forum

Ambiguous & Silent

Ambiguous

Mini case card

id	AT-StiAG_GoettweigOSB_1483_XII_01_tenor.txt
modernPlaceName	Graz
flame_cluster_id	SINGLETON
cluster_size	1
in_atlas_1km	NO
in_atlas_5km	YES
Urbanity Level	ambiguous
category	C_atlas_not_urban

“unnsern burgern zu Stain”

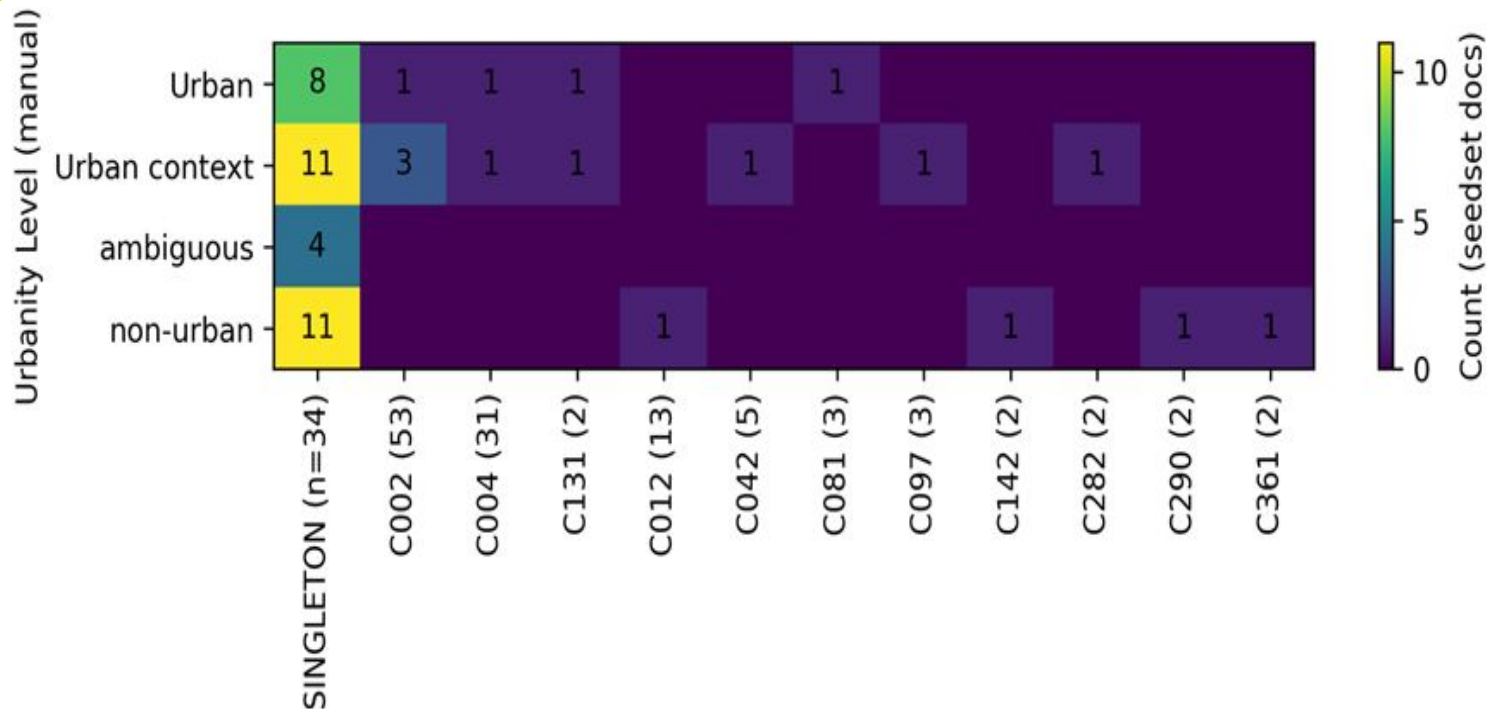
Silent

Mini case card

id	AT-StiAH_HeiligenkreuzOCist_1443_II_28_tenor.txt
modernPlaceName	—
flame_cluster_id	SINGLETON
cluster_size	1
in_atlas_1km	NO
in_atlas_5km	NO
Urbanity Level	Urban
category	B_urban_not_atlas

des rats der stat zu Wienn”

Output



FLAME cluster (total corpus size in parentheses; SINGLETON shows n in seedset)

Distribution of manual urbanity levels across FLAME similarity clusters in the seedset. Cluster labels indicate total cluster size in the full corpus; cell values show seedset counts.

Output(2)

Map A2 — Controlled overlay (seedset Urbanity Level)

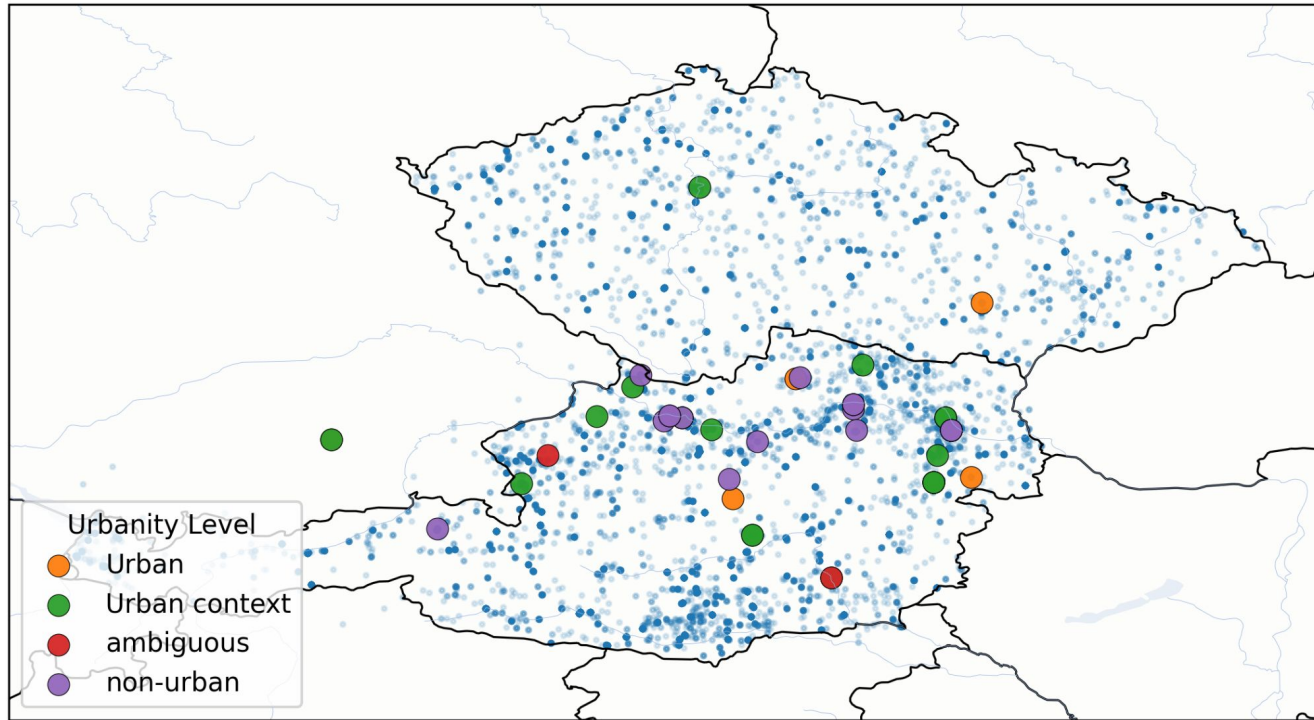


Fig.2 Overlay of the manually classified seedset (50 documents) on the baseline place-of-issuing map. Large points represent controlled cases (Urban, Urban context, Ambiguous, Non-urban), while the background shows the full spatial distribution.

Re-usable workflow (7 steps)

1

Preprocessing
Corpus

2

Information
extraction
(atomid/id/placeNa
me)

3

Similarity +
clustering
(FLAME)

4

External spatial
signals
(atlas 1km/5km)

5

Seedset +
classification criteria
(Indicators/Confunde
r)

6

Intersect layers
(intersection table)

7

Case review
+ update
classification
criteria

Conclusion

1) Map ≠ label

Maps help explore place-of-issuing, but do not define urbanity.

2) Cluster ≠ city

FLAME clusters organize textual similarity, not urban institutions.

3) Control via intersection

Urbanity is controlled by intersecting external signals + diplomatic reading.



<https://www.didip.eu>
@DiDip_ERC@fedihum.org

UNIVERSITY OF GRAZ

Department of
Digital Humanities



Thanks!

Workshop - Digital Methods in Diplomatic Research

Giuseppe Consolo - University of Graz - 13/02/2026
giuseppe.consolo@uni-graz.at