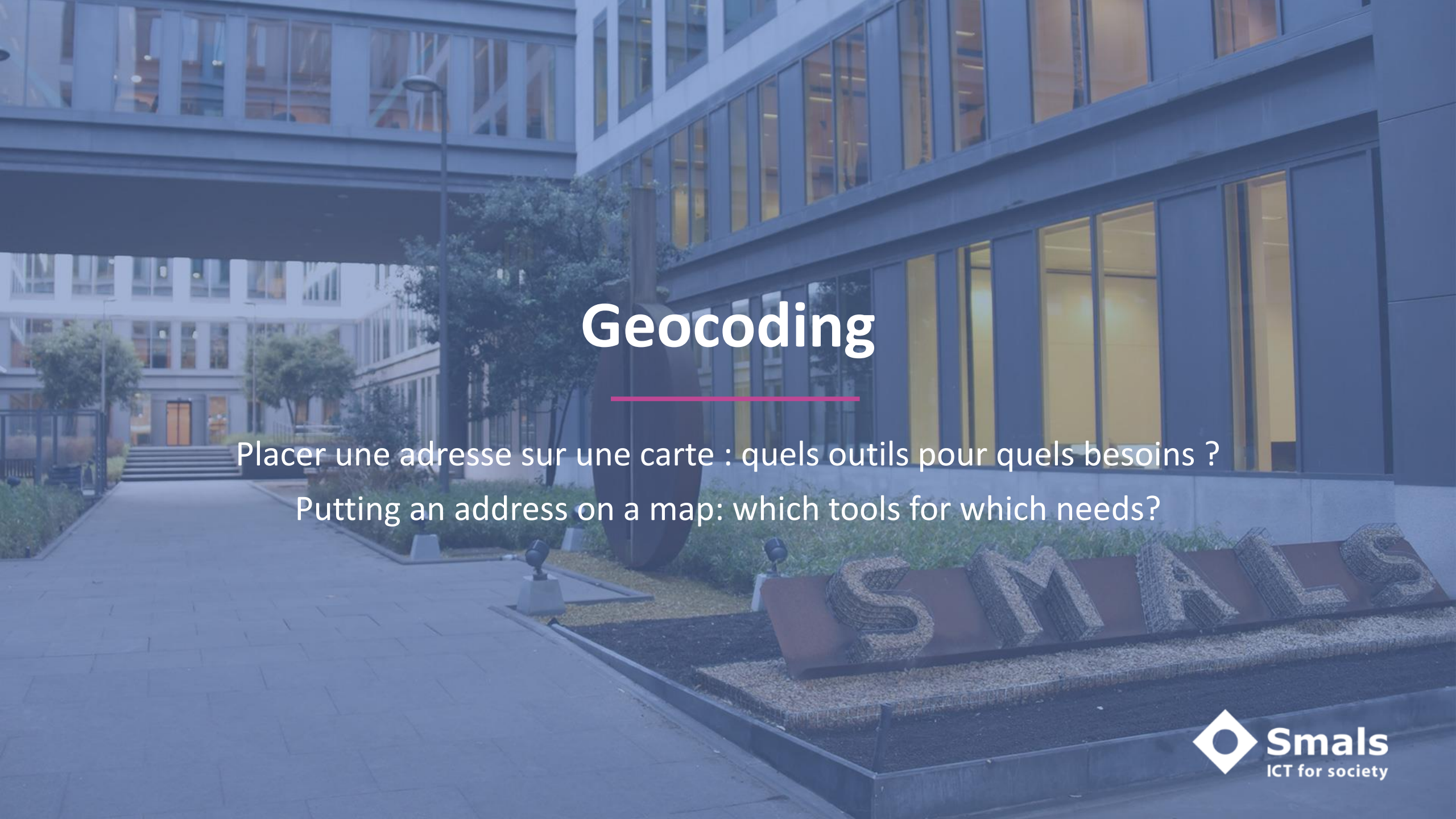


Geocoding

Placer une adresse sur une carte : quels outils pour quels besoins ?

Putting an address on a map: which tools for which needs?

Smals
ICT for society

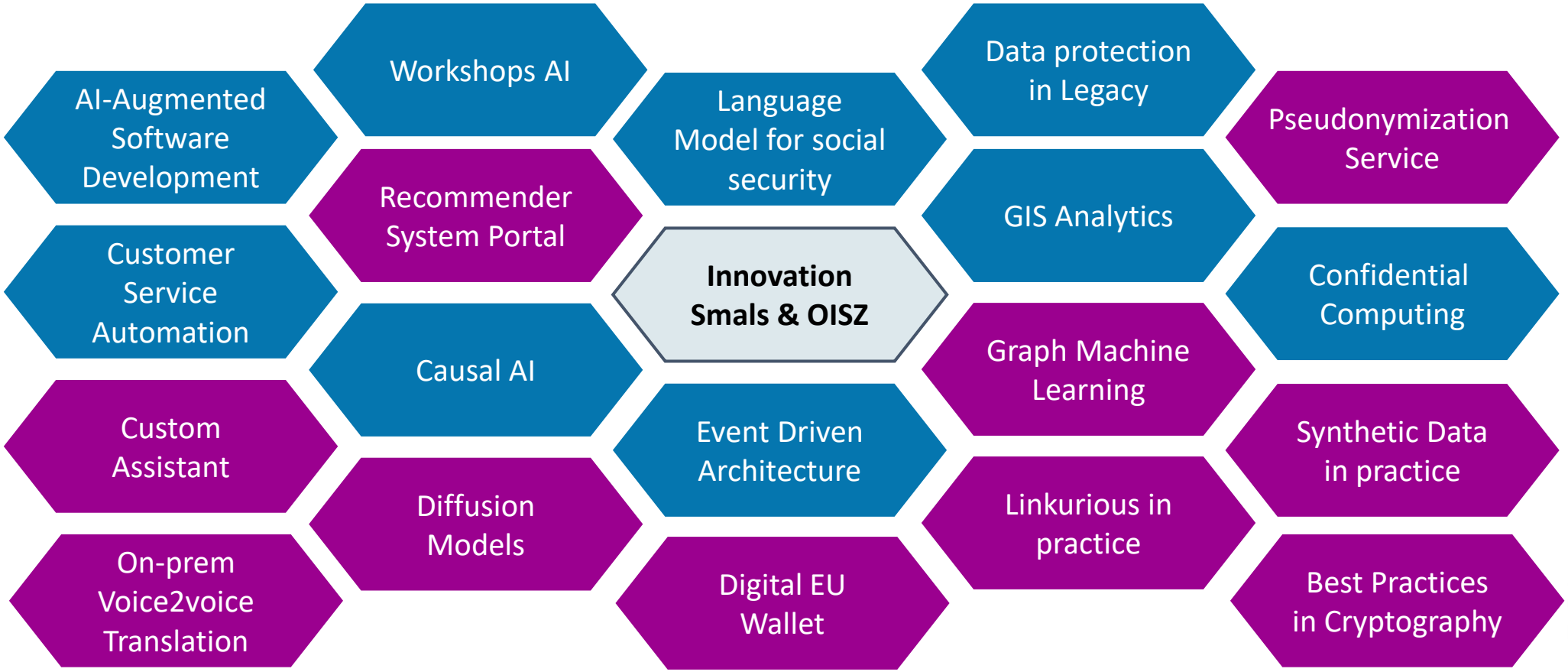
Smals Research 2023

Innovation with new technologies

Consultancy & expertise

Internal & external knowledge transfer

Support for going live



Smals
ICT for society

www.smalsresearch.be

Agenda

- Concepts
- Evaluation
- Nominatim based solution
- BestAddress+Nominatim based solution
- Pelias based solution
- Conclusion

Map: OpenStreetMap

Smals
ICT for society

Agenda

- **Concepts**
  - **Definitions**
  - **Challenges**
  - **Tools**
- Evaluation
- Nominatim based solution
- BestAddress+Nominatim based solution
- Pelias based solution
- Conclusion

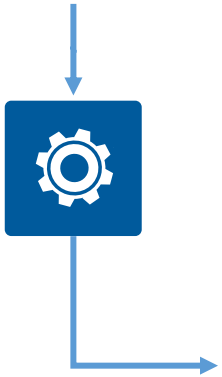
Map: CadGIS

Smals
ICT for society

Geocoding: concept

Geocoding = (1) Address standardization + (2) Geolocation
Geocoder = API (Application Programming Interface)

"Av. Fonsny 20, 1060 Bruxelles"



```
{"street"  : "Avenue Fonsny",
 "housenbr": "20",
 "zipcode" : "1060",
 "city"    : "Saint-Gilles",
 "country" : "Belgique",
 "location": [50.8358,4.3361]}
```

Or polygon

(1) Address standardization:

- Data quality:
  - Standardization
  - Deduplication
  - Matching
- Auto-completion (type-ahead)
- …

(2) Geolocation:

- Visualization
- Controls
- Spatial (GIS) analytics
- …

Smals
ICT for society

Geocoding: concept

Address standardization = Mapping
Geolocation = « side effect »/bonus

"Av. Fonsny 20, 1060 Bruxelles"

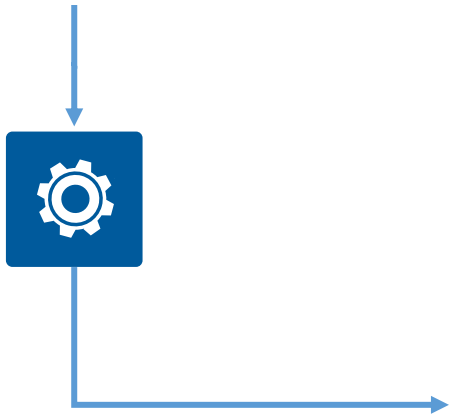
Street	Number	Zipcode	Locality	Location
…	…	…	…	…
Avenue Fonsny	20	1060	Saint-Gilles	50.8358, 4.3361
Avenue Fonsny	22	1060	Saint-Gilles	50.8359, 4.3358
…	…	…	…	…

```
{"street"   : "Avenue Fonsny",
 "housenbr": "20",
 "zipcode"  : "1060",
 "city"     : "Saint-Gilles",
 "country"  : "Belgique",
 "location": [50.8358,4.3361]}
```

Smals
ICT for society

Reverse geocoding

```
lat: 50.8358, lon: 4.3362, range: 10m
```



```
{“street”   : “Avenue Fonsny”,
 “housenbr”: “20”,
 “zipcode” : “1060”,
 “city”     : “Saint-Gilles”,
 “country” : “Belgique”,
 “location”: [50.8358,4.3361]}
```

Smals
ICT for society

Challenge 1: parsing

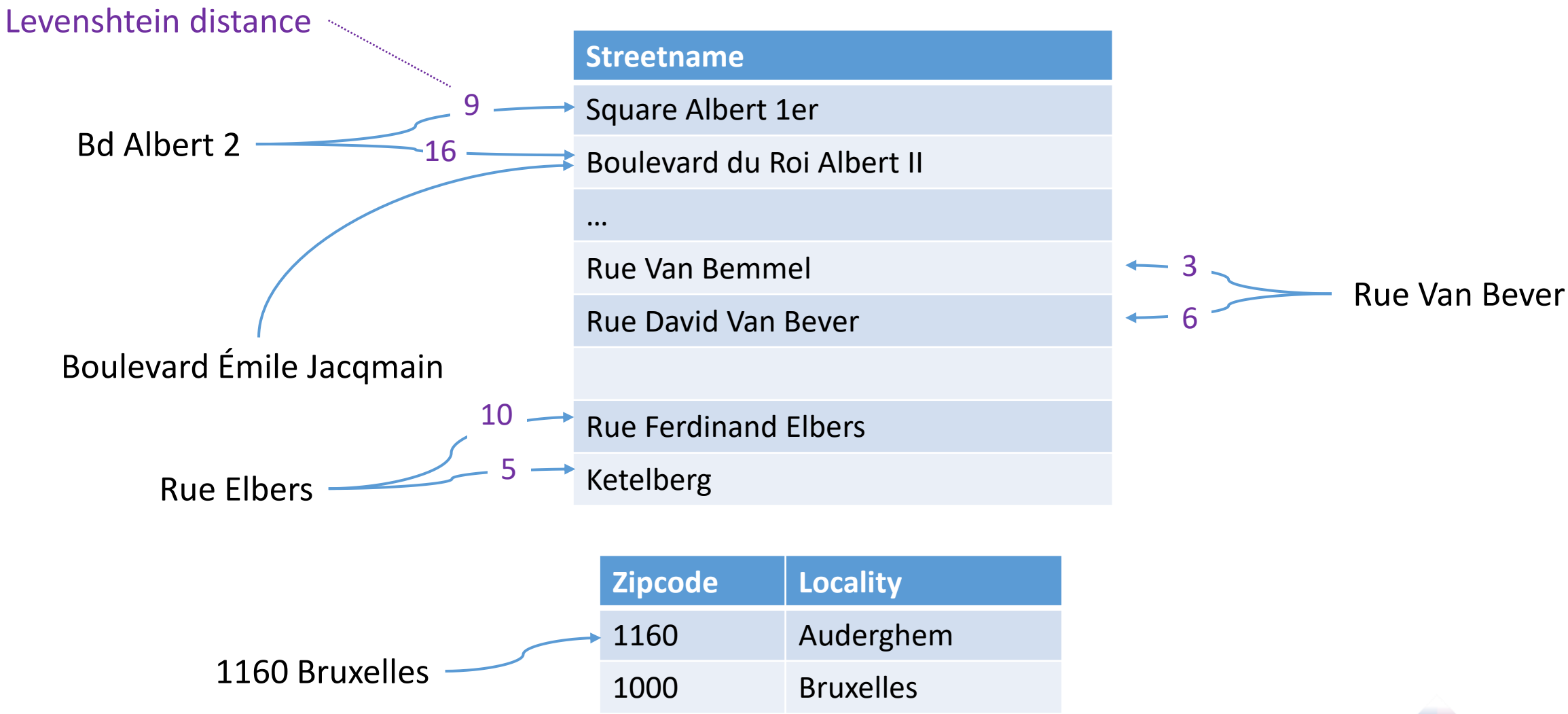
“Av. Fonsny 20, 1060 Bruxelles” ➔

```
{“street”  : “Av. Fonsny”,
 “housenbr”: “20”,
 “zipcode” : “1060”,
 “city”    : “Bruxelles”}
```

- Very complex problem looking very simple!
  - Not always « Avenue Fonsny 20, 1060 Saint-Gilles »
  - « 1600/2 Chée de Wavre, 1160 Bruxelles »: easy for a human, not for a computer
- Box information : “10 A”:
  - “housenumber”: “10 A”,
  - or “housenumber”: “10”, “box”: ”A” (10/A, 10 bt A, 10 apt A…)
- Addresses often contain extra/useless info:
  - Smals, Avenue Fonsny…
  - Avenue Fonsny 20 (Saint-Gilles), 1060 Bruxelles
  - « Sonnez ‘Tartempion’ », « Au dessus de la poste »…

Smals ICT for society

Challenge 2: matching



Many quasi-homonyms ; exm in BRU: Lapins vs Lupins vs Lutins, Augustins vs Augustines, Savoir vs Lavoir, Cologne vs Pologne

Tools

- Cloud commercial providers:
  - Here we go
  - Bing maps
  - Google maps
  - TomTom
- Sometimes free in specific conditions

- Mainly 3 open-source/open-data solutions:
  - OpenStreetMap/Nominatim
  - Photon (OSM + ElasticSearch)
  - Pelias
- On-premise possible for a specific region

Smals
ICT for society

Partial geocoders (Belgium)

Geocoder = API - standardization - geolocation

- Trillium @ Smals (Data Quality tool):
  - API: OK ; Standardization: OK (for millions of records)
  - Geolocation: NO

- BeST (BestAddress/BOSA) API:
  - API: OK ; Geolocation: OK
  - Standardization: (almost) no cleansing (assumes correct addresses)

- PhacochR (Paradigm (ex CIRB/CIGB) + IGEAT (ULB)):
  - Standardization: OK ; Geolocation: OK
  - API: NO, only batch

Cloud vs On-premise : pros/cons

	Cloud solutions	On-premise solutions
Setup complexity	☺	☹
Infra costs	☺	☹
GDPR, Confidentiality	☹	☺
Cost for large volumes (time, €)	☹	☺
Internet access	☹	☺
Quality	☺	😐
Worldwide coverage	☺	☹

Quality indicators

- In most cases, geocoding response is a JSON (or XML) document
- Beside geocoded address, geocoders (often) provide a quality indicator
- Mainly 2 ways of scoring response quality:
  - Precision: building, interpolated, street, city, country…
  - Confidence level

Smals
ICT for society

Quality indicator: precision

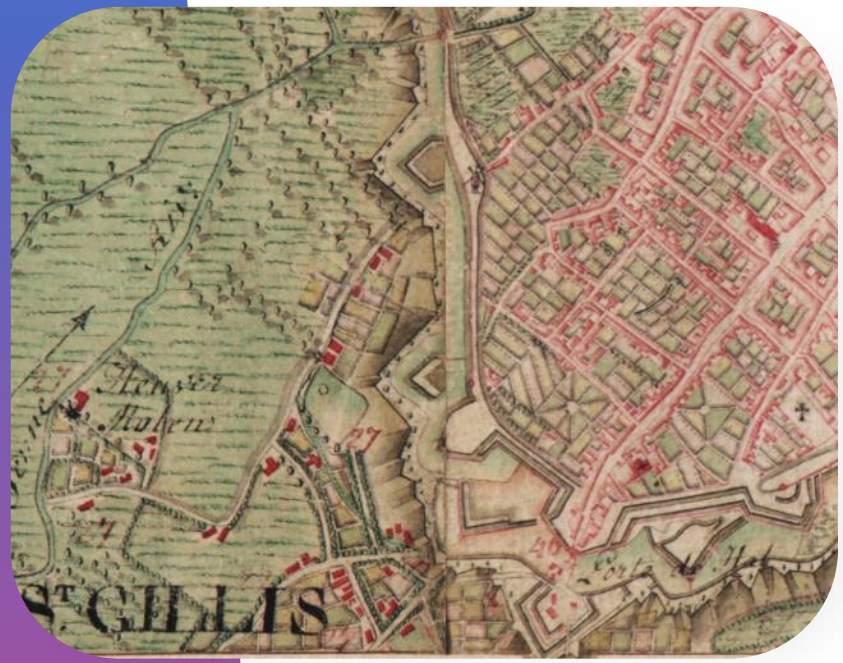
Geocoder	Indicator	Building	Street	City
Google	geometry > location_type	ROOFTOP RANGE_INTERPOLATED	GEOMETRIC_CENTER	APPROXIMATE
Here	MatchLevel	houseNumber	street	city postalCode
Bing	entityType	Address	RoadBlock	PopulatedPlace Postcode1
TomTom	type	Point Address	Street Address Range Cross Street	Geography
Nominatim	class	building amenity shop place	highway	
	place_rank	30	26-27	14-25

Quality indicator: confidence

Geocoder	Indicator	Range
Here	Relevance	0-1 (% of input matching the result)
	MatchQuality	0-1 for Street, HouseNumber, City (attribute level relevance)
	MatchType	pointAddress or interpolated
TomTom	matchConfidence	0-1
Bing	confidence	“High”, “Medium”, “Low”
	matchCode	list of “Good”, “Ambiguous”, “UpHierarchy”

Smals ICT for society

Agenda

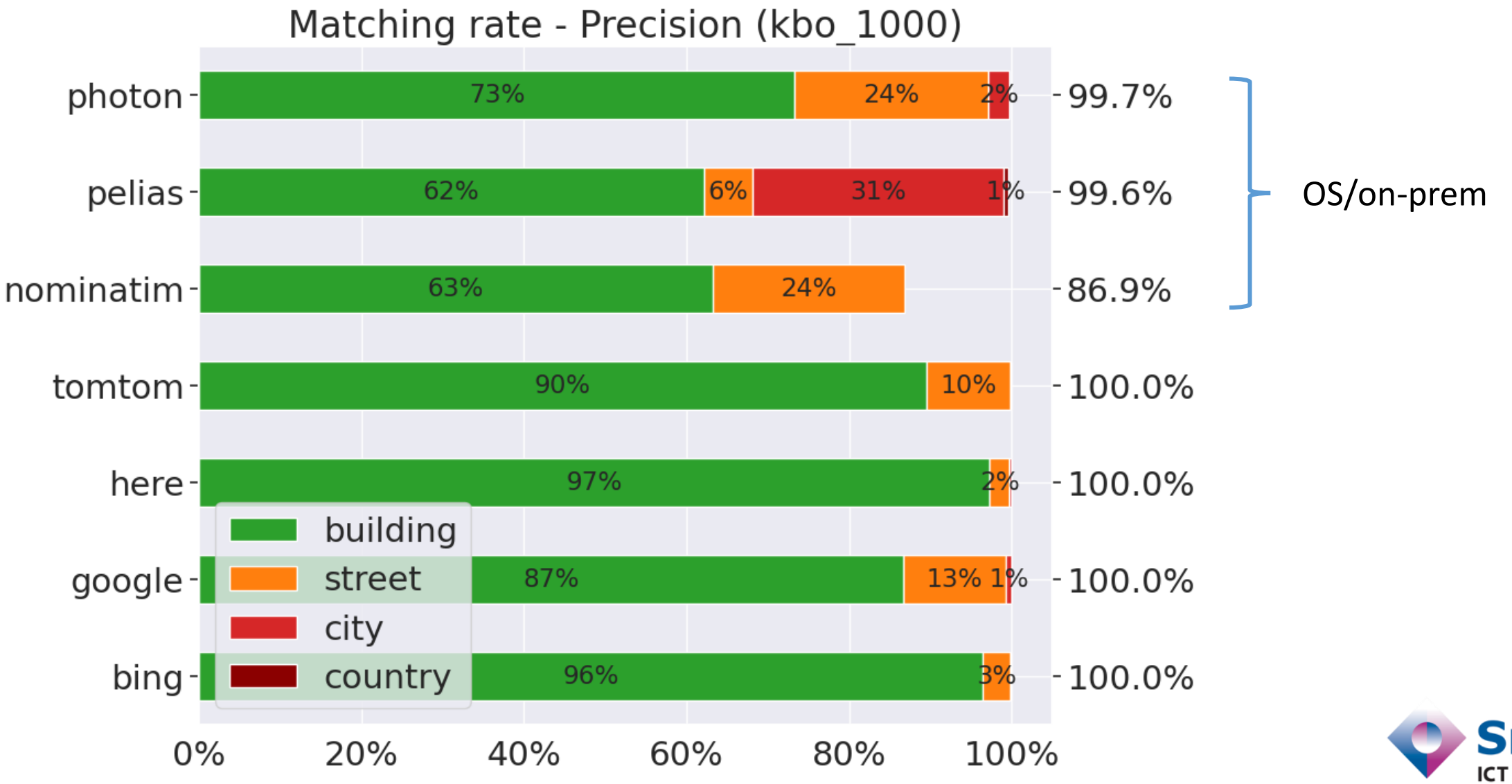


- Concepts
- **Evaluation**
  - **Matching rate**
  - **Median point methodology**
  - **Duration**
  - **Pricing**
- Nominatim based solution
- BestAddress+Nominatim based solution
- Pelias based solution
- Conclusion

Map: Ferraris (KBR)

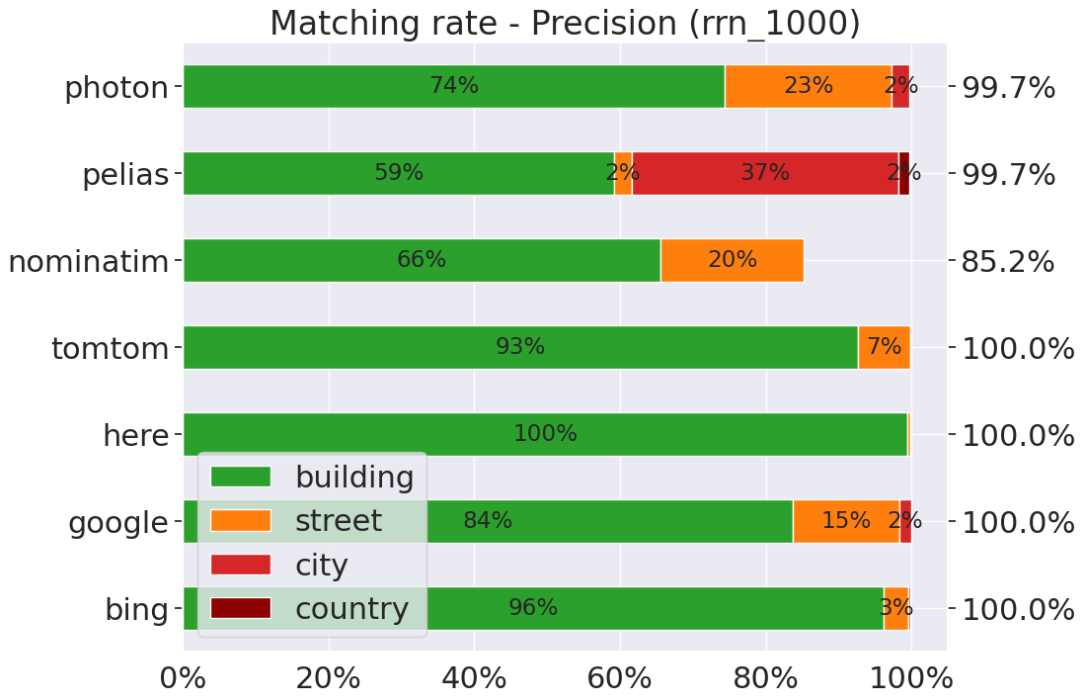
Smals
ICT for society

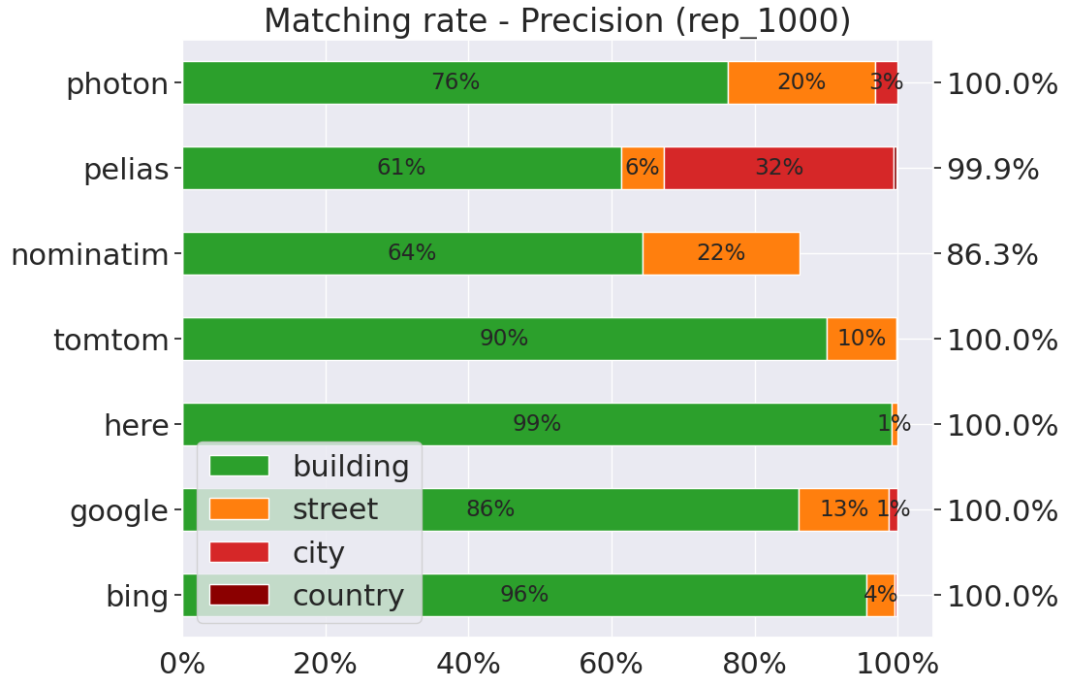
Matching rate

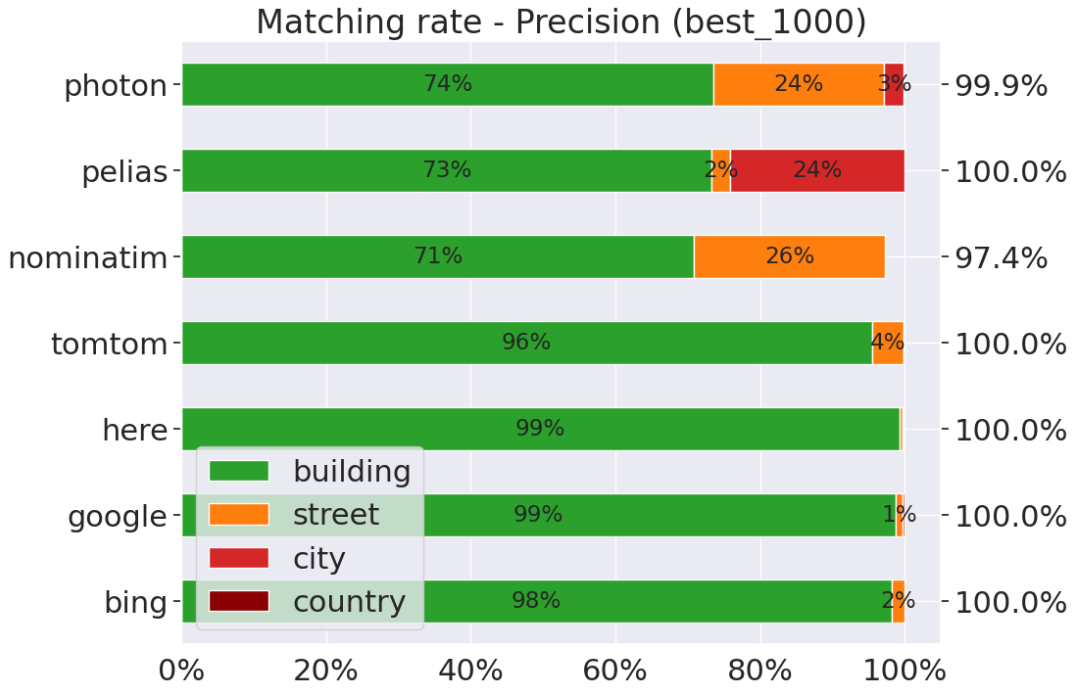


Dataset : 1 000 (Belgian) addresses from KBO OpenData

Matching rate







RRN: National Register (IBZ)
REP: Employers repertorium (ONSS/RSZ)
KBO: Crossroads Bank for Enterprises (SPF eco)
BeST: BestAddresses (BOSA)

Performances

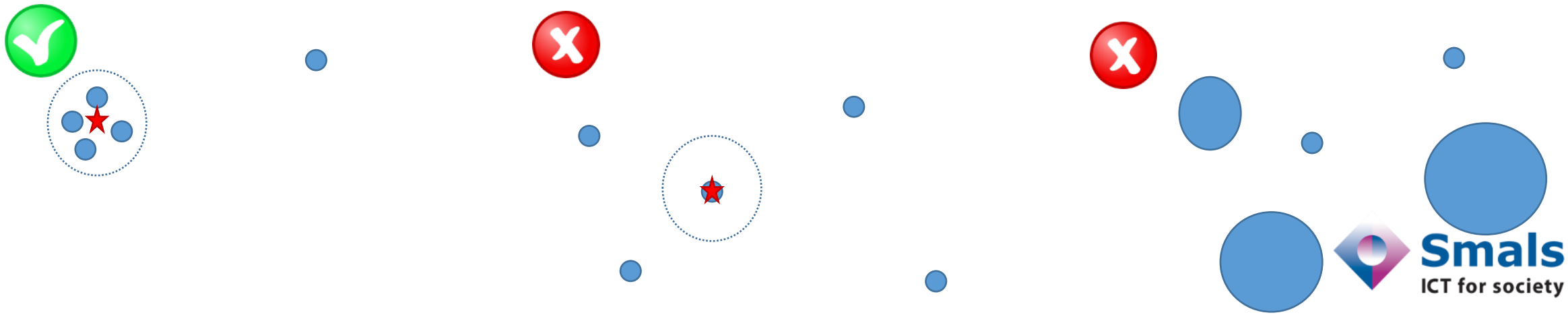
- Giving an answer does not make this answer to be correct!
- Ideal situation: testing against a reference address list. But:
  - Building such a list is complex!
  - List will be (intrinsically) biased

→ We propose a methodology based on a “majority vote”

Smals
ICT for society

Methodology

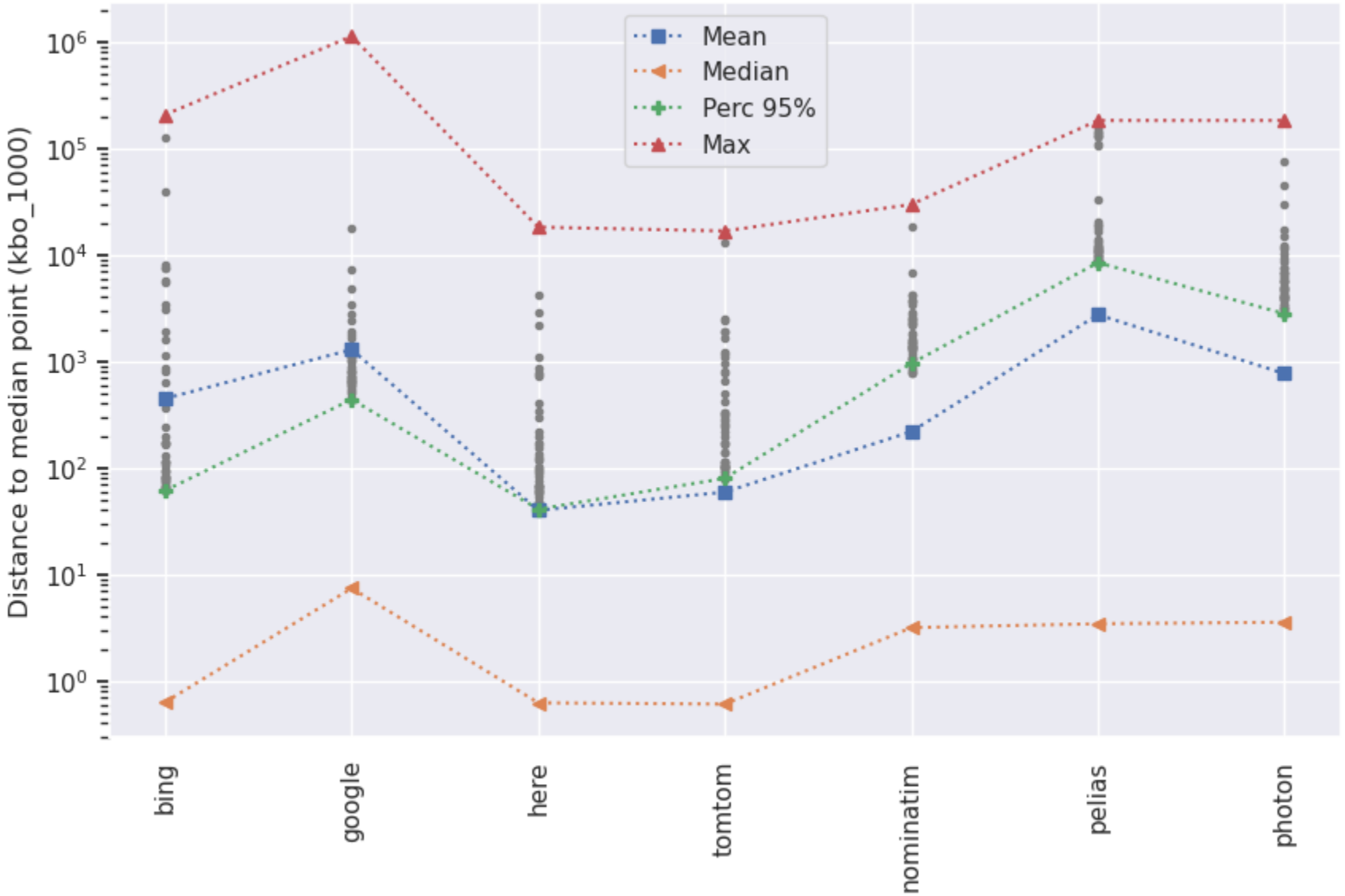
- Collect a (not too clean) address dataset → 1 000 addresses from KBO
- Send this dataset to several coders → here, bing, nominatim, google…
- Use median point for each address as reference, with some conditions:
  - Enough (precise) matches → 957 with ≥ 3/7 “building” matches
  - Enough answers close to median → 932 with ≥ 3/7 within 100 m
- Compute misc. stats about « distance to median point »



Smals
ICT for society

Distance to median point

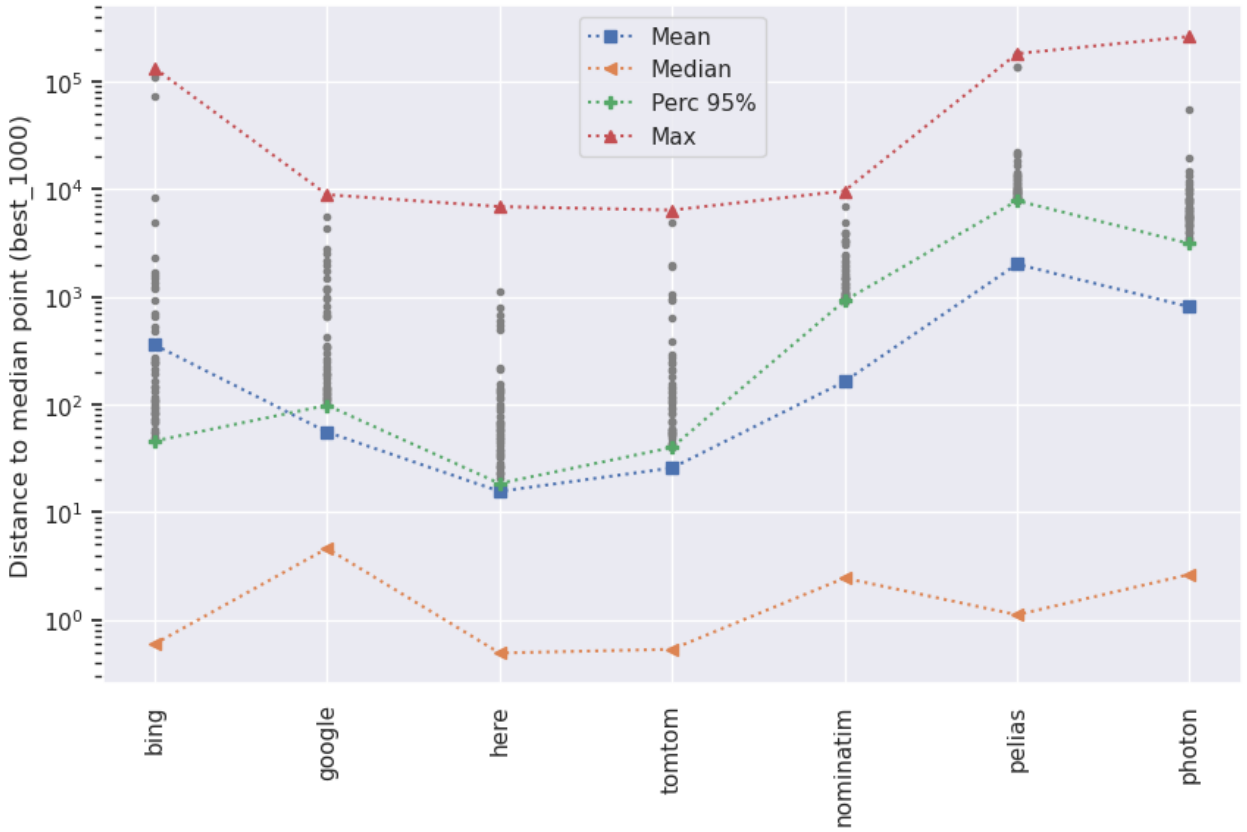
OS/on-prem

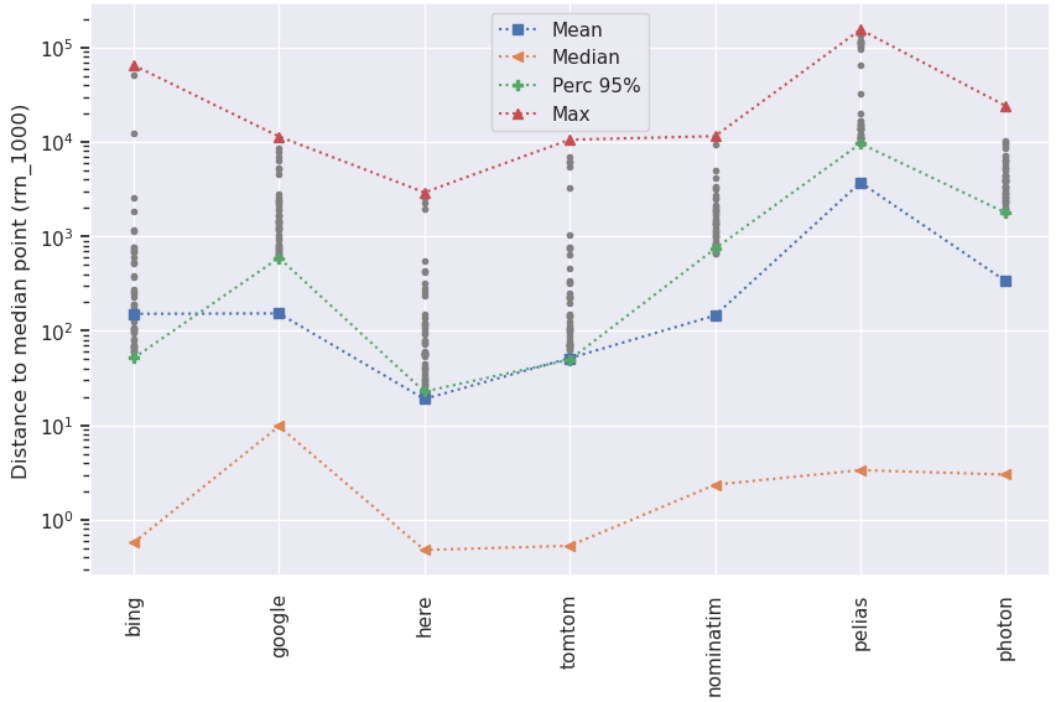


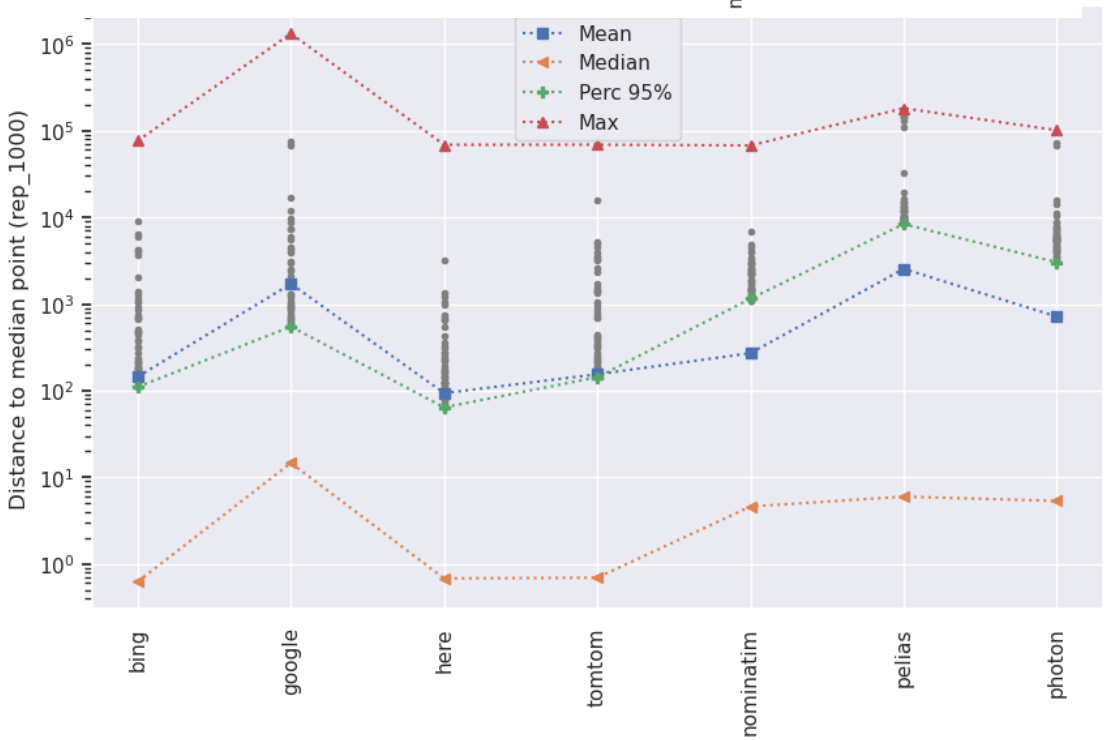
The lower the better!

Smals
ICT for society

Distance to median point

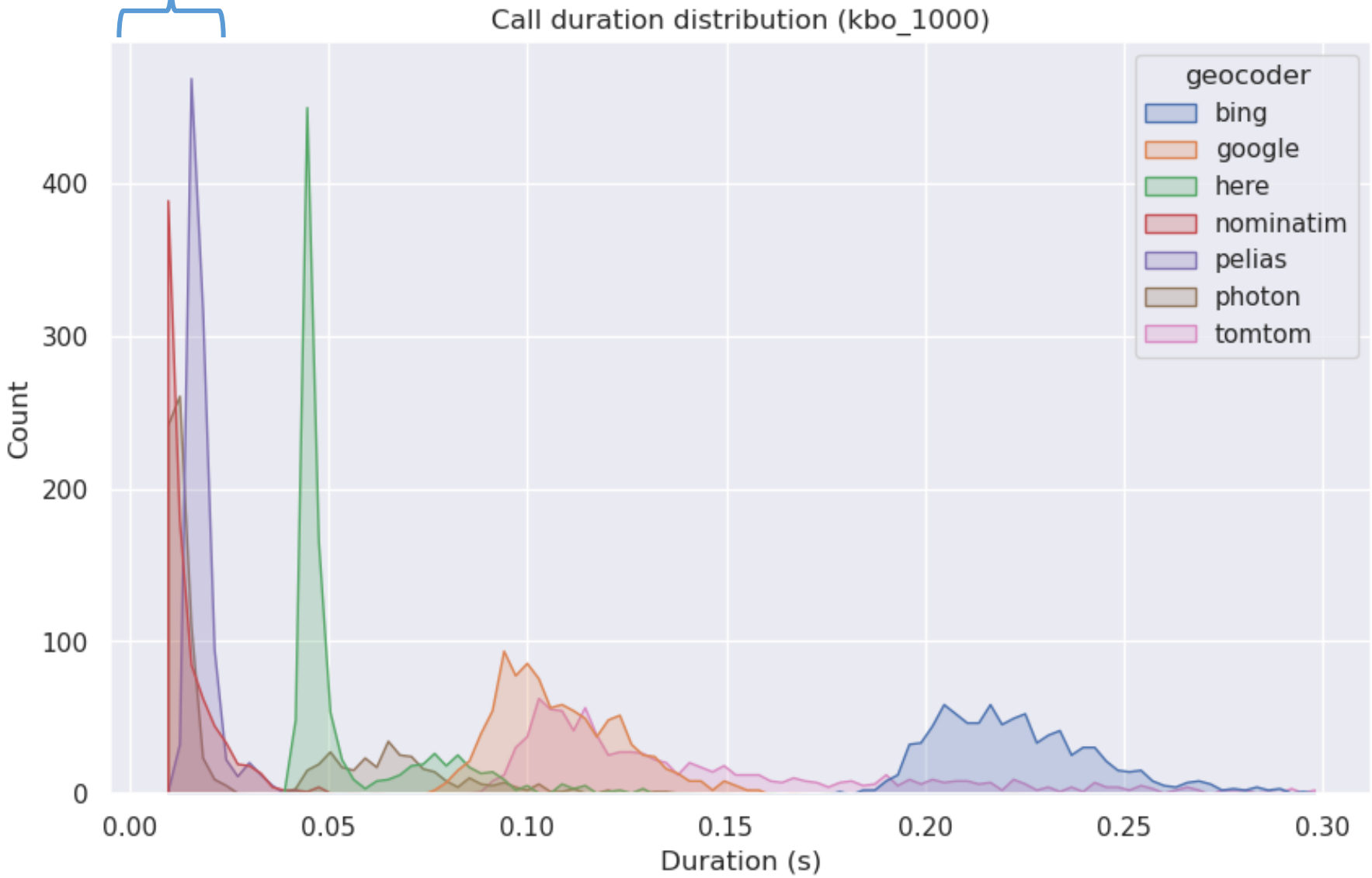






Call duration

OS/on-prem (→ same machine!)



Smals
ICT for society

Pricing: free usage plan

Geocoder	Free requests	Per day
Google	40 000 /month	1 333
Here	30 000 /month	1 000
Tomtom	2 500 /day	2 500
Bing	125 000 /year	342
MapBox	100 000 /month	3 333

Google: credit card required!

Smals
ICT for society

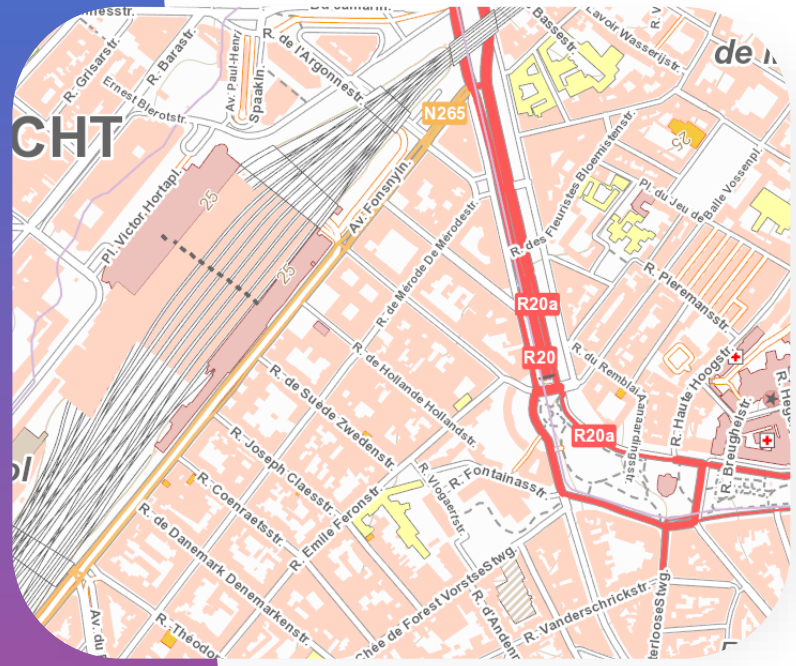
Pricing: paid usage

Geocoder	Price min (€/1000)*	Price max
Google	3.70	4.60
Here	0.53	0.66
Tomtom	0.50	0.50
Bing	Contact sales?	
MapBox	0.42	0.69

* Without contacting sales
  1$ = 0.93 €

Smals ICT for society

Agenda

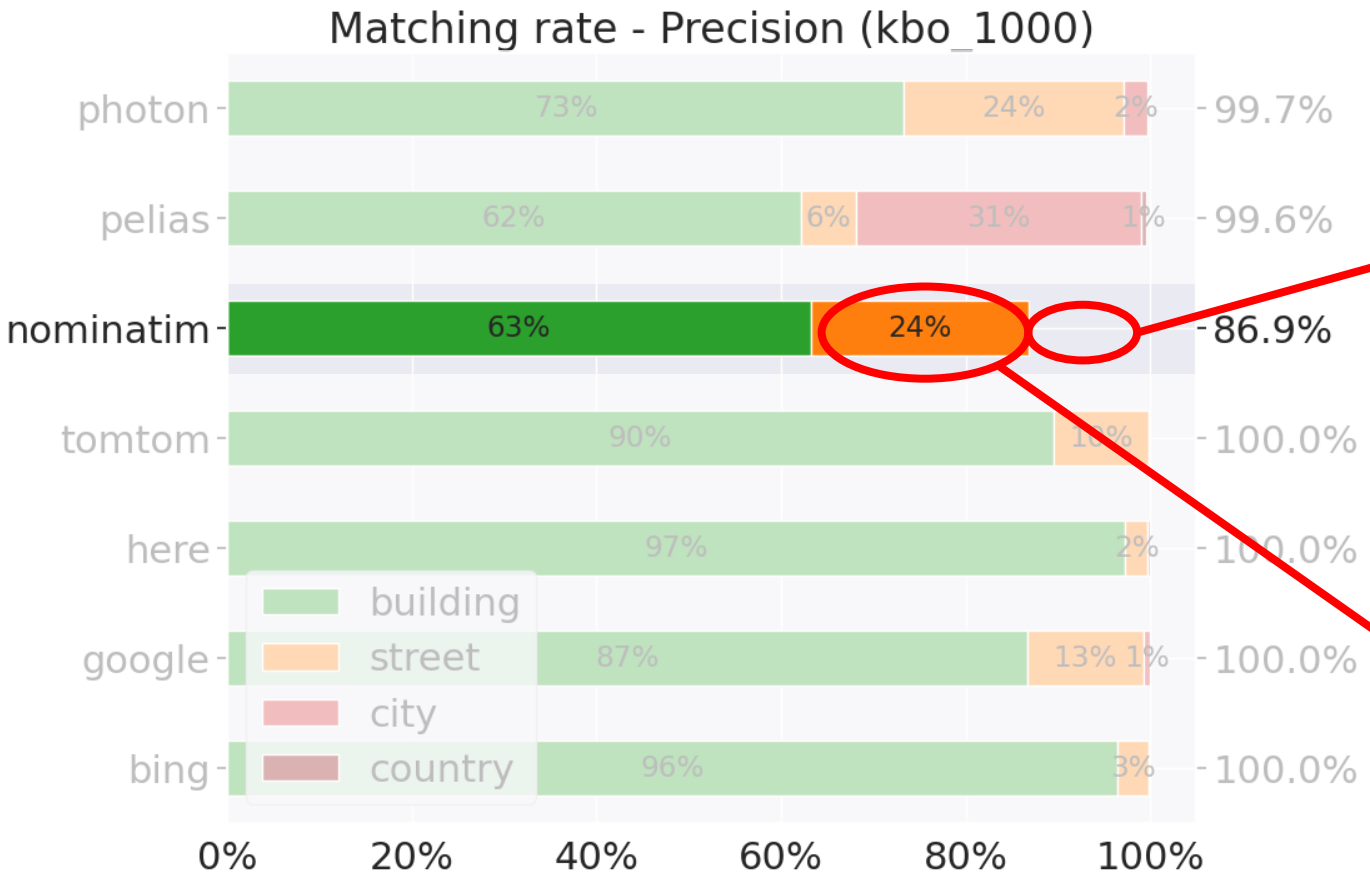


- Concepts
- Evaluation
- **Nominatim based solution**
  - **Nominatim weaknesses**
  - **NominatimWrapper solution**
- BestAddress+Nominatim based solution
- Pelias based solution
- Conclusion

Map: geo.be (IGN-NGI)

Smals ICT for society

Weaknesses of Nominatim/OSM



Not very **robust**

Often **street level** → not a problem for many applications?

Weak output **coherence** (~10 fields for ”street”)

Smals
ICT for society

Nominatim: weak robustness (input)

- In general, does not support “small typos”
- Robustness for housenumbers:
  - Avenue Fonsny 20, 1060 Bruxelles → OK
  - Avenue Fonsny 20 bte 2, 1060 Bruxelles → No result
  - Avenue Fonsny 20/2, 1060 Bruxelles → Street level (skip 20/2)
- Robnustness with names:
  - Avenue Fonsny 20, 1060 Bruxelle → No result
  - Avenue Fonsni 20, 1060 Bruxelles → No result
  - Avenu Fonsny 20, 1060 Bruxelles → No result
- With an unknown housenumber:
  - Rue des Guildes, 22, 1000 Bruxelles → Street level (skip 22)
  - Rue Gray 96, 1040 Etterbeek → No result
  - … but : Rue Gray, 1040 Etterbeek → Street level
- Real life examples : « Avenue Fonsny, en face du 20 », « … (entrée rue X) », « Galerie Y »

Smals
ICT for society

Nominatim: weak coherence (output)

https://nominatim.openstreetmap.org/search.php?q=**Avenue+Fonsny+20,+1060+Bruxelles**&format=jsonv2&addressdetails=1

➔ { place_id: 48588284,

lat: "50.8358216", lon: "4.3386884", (…)

address: {

house_number: "20",

road: "Avenue Fonsny",

town: "Saint-Gilles", (…) }

road, pedestrian, footway, cycleway, path, address27, construction, hamlet, park, square

town, village, city, city_district, county

Smals ICT for society

Nominatim: observations

- Often, simple change allows matching:
  - Av. Fonsny → Avenue Fonsny
  - Avenue Fonsny 20 bte 2 → Avenue Fonsny 20
  - Rue Gray 96 → Rue Gray
  - Rue de Namur, Spy (Jemeppe-s-s) → Rue de Namur, Spy
- Simple tools allow to “clean” addresses:
  - Regex: “^av[\.]” ➔ “avenue ”
  - Libpostal (address parser with ML) :

“Smals, Av Fonsny 20 bte 5, 1060 St-Gilles” →

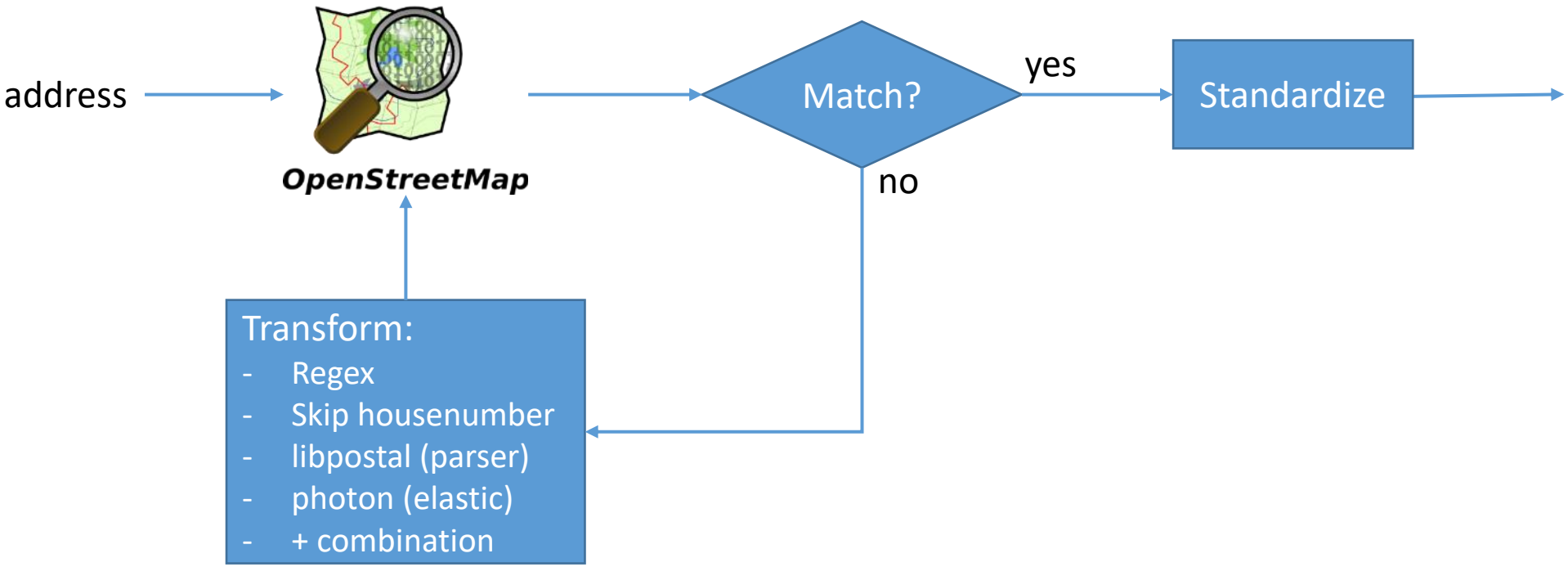
```
[('smals',       'house'),
 ('av. fonsny',  'road'),
 ('20',          'house_number'),
 ('bte 5',       'unit'),
 ('1060',        'postcode'),
 ('st-gilles',   'city')]
```

  - Photon = Nominatim+ElasticSearch

Smals
ICT for society

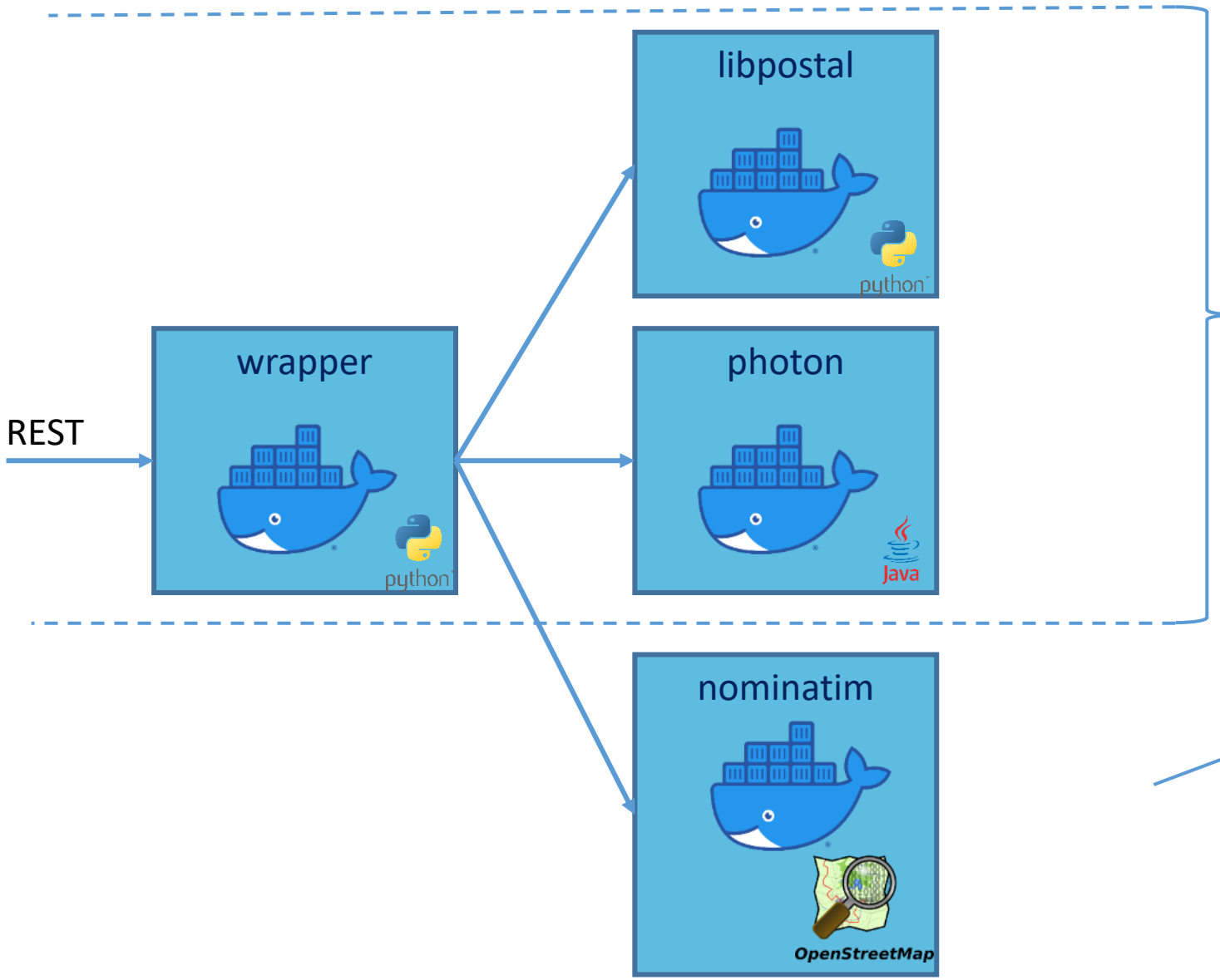
NominatimWrapper

- To address the robustness and output coherence problem, we build a (REST API) « wrapper » around Nominatim
- Main idea: if Nominatim does not recognize an address, « clean it » beforehand



Smals
ICT for society

Architecture



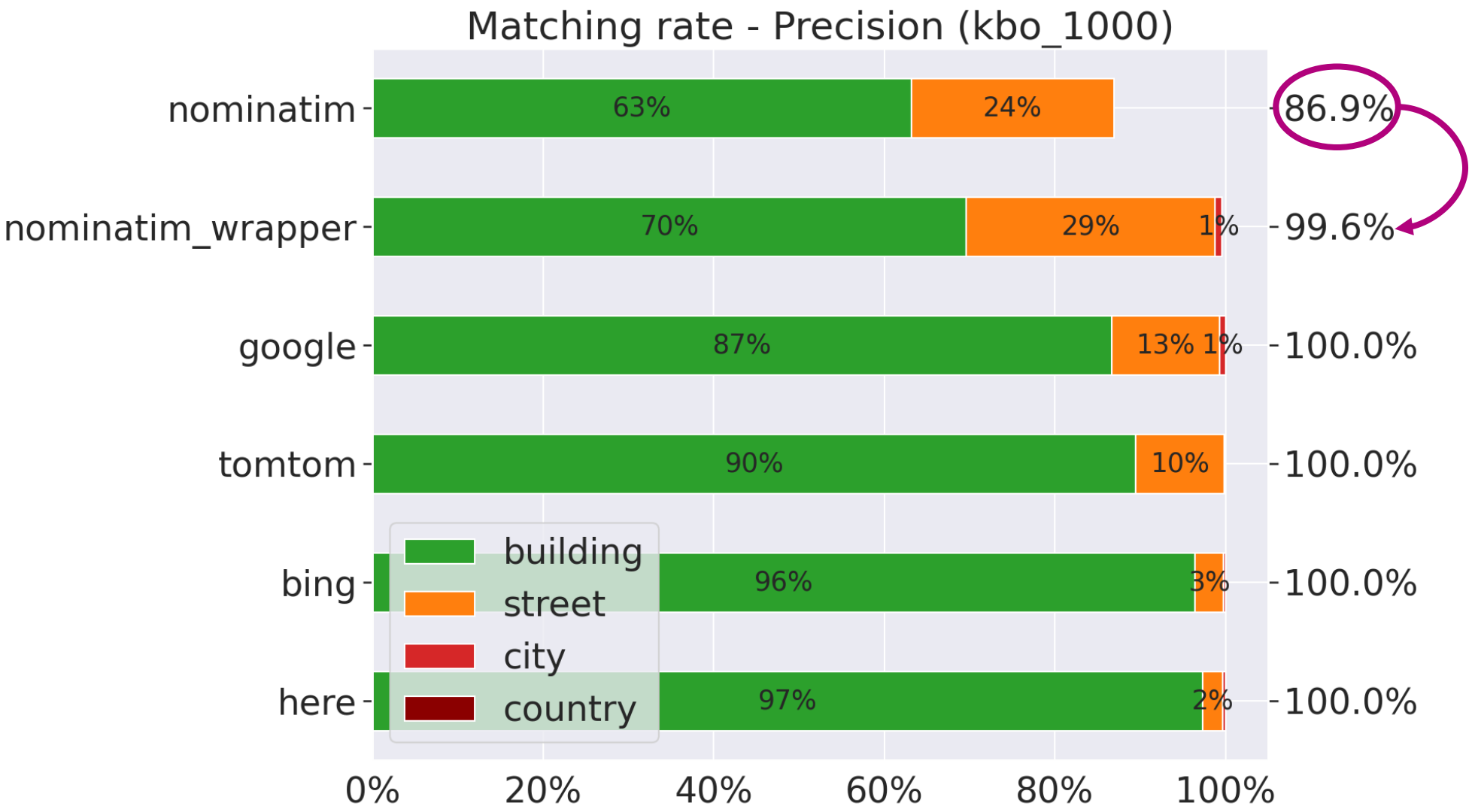
NominatimWrapper
(docker-compose)

mediagis/nominatim
Options:
- Region
- Level (admin/street/address/full)

Belgium + address: ~ 30GB

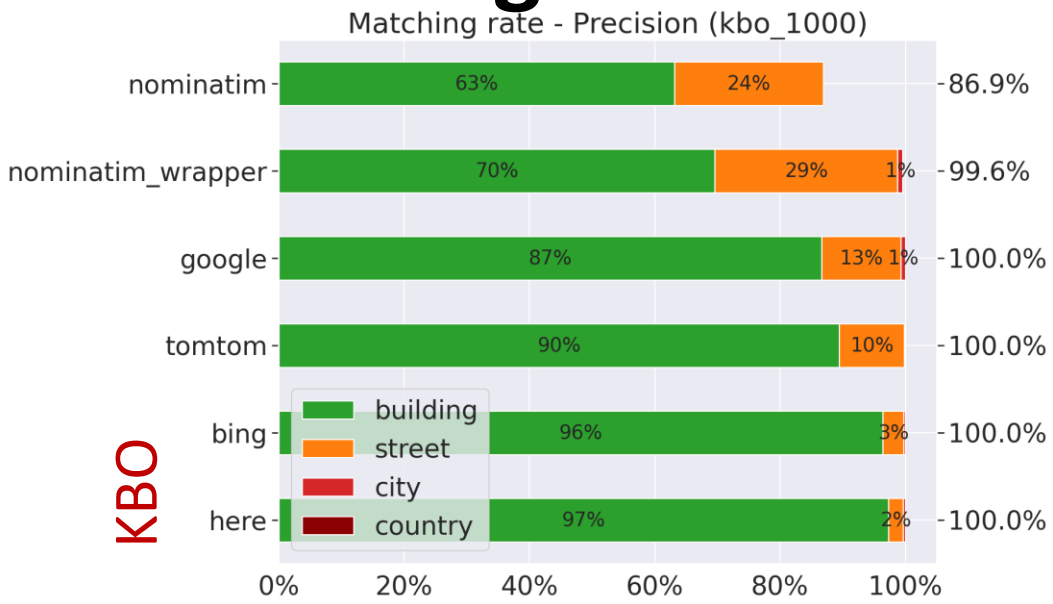
Smals
ICT for society

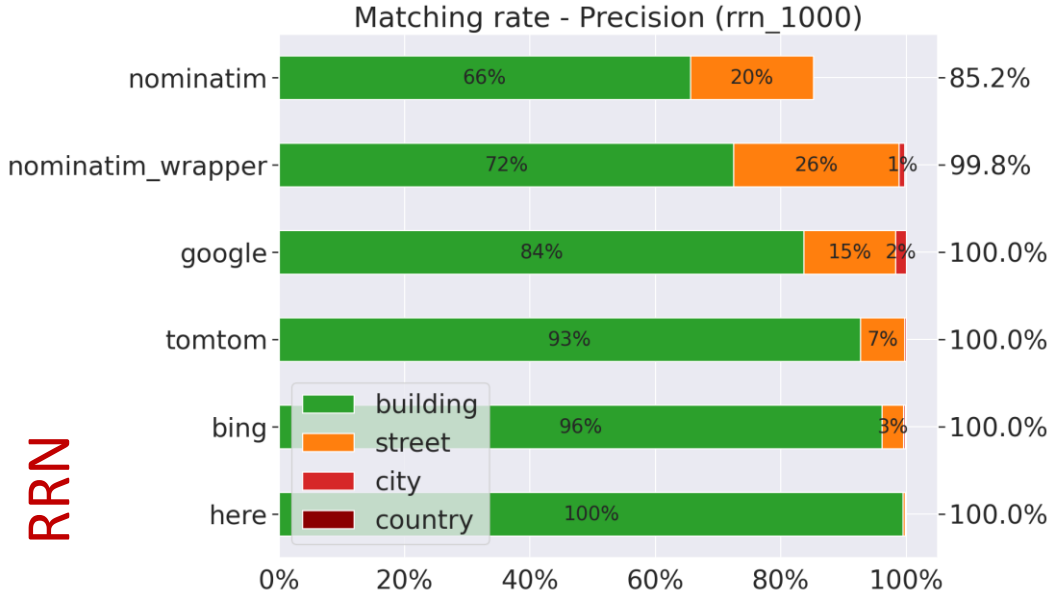
Matching rate

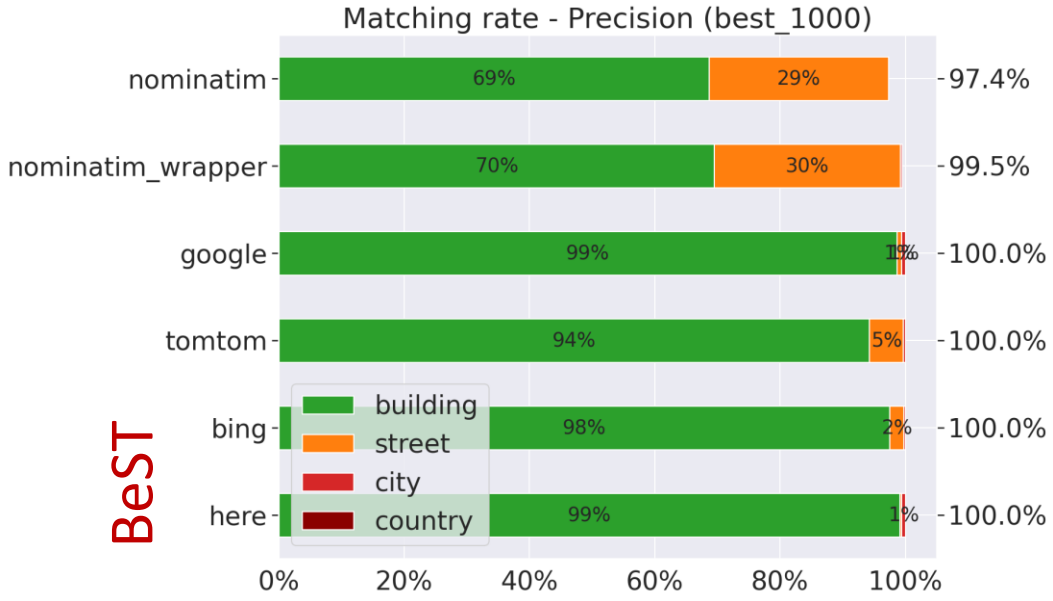


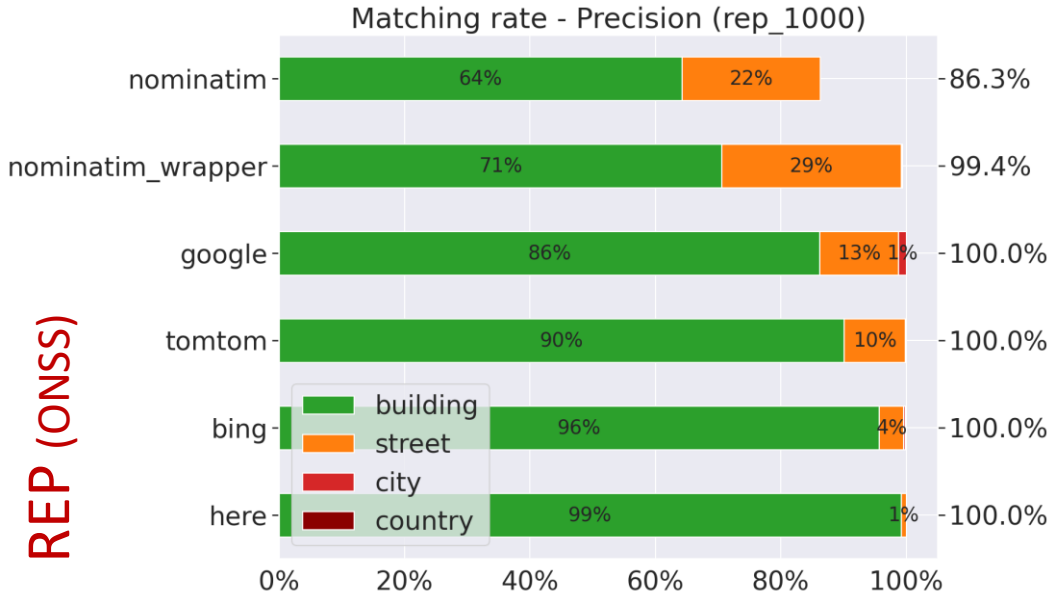
Smals
ICT for society

Matching rate









References

- Published as Open Source (Github). Deployement in two steps:
  - Nominatim: 1 Docker command line
  - NominatimWrapper: config + 1 Docker command line
  - One script to rule them all
- In ReUse Catalog: https://www.ict-reuse.be/fr/service/geocodage-avec-nominatim-ameliore
- References:
  - Github: https://github.com/SmalsResearch/NominatimWrapper/
  - Swagger GUI: https://nominatimwrapper.smalsrech.be/doc
  - Blog: https://www.smalsresearch.be/tag/geocoding/

NominatimWrapper and Trillium

- Integrated in Trillium batch
- Used for foreign addresses in a project for INASTI/RSVZ
- API integration in progress

TRILLIUM SOFTWARE

Smals ICT for society

NominatimWrapper: conclusions

- On-premise solution for geocoding
- Built on open-source soft/data (Nominatim, libpostal, photon)
- Match rate: ~identical to commercial solutions (from ~85 to 99,5 %)
- Precision: building or street level for ~99 % of addresses (enough for most of our cases)
- Also available for foreign countries (with a structure street/postcode/city)
- In ReUse Catalog since June 2023

Smals
ICT for society

Agenda

- Concepts
- Evaluation
- Nominatim based solution
- **BestAddress+Nominatim based solution**
  - **BestAddress**
  - **BestAddres+NominatimWrapper**
- Pelias based solution
- Conclusion

Map: bruciel.brussels

Smals
ICT for society

BeSt Addresses

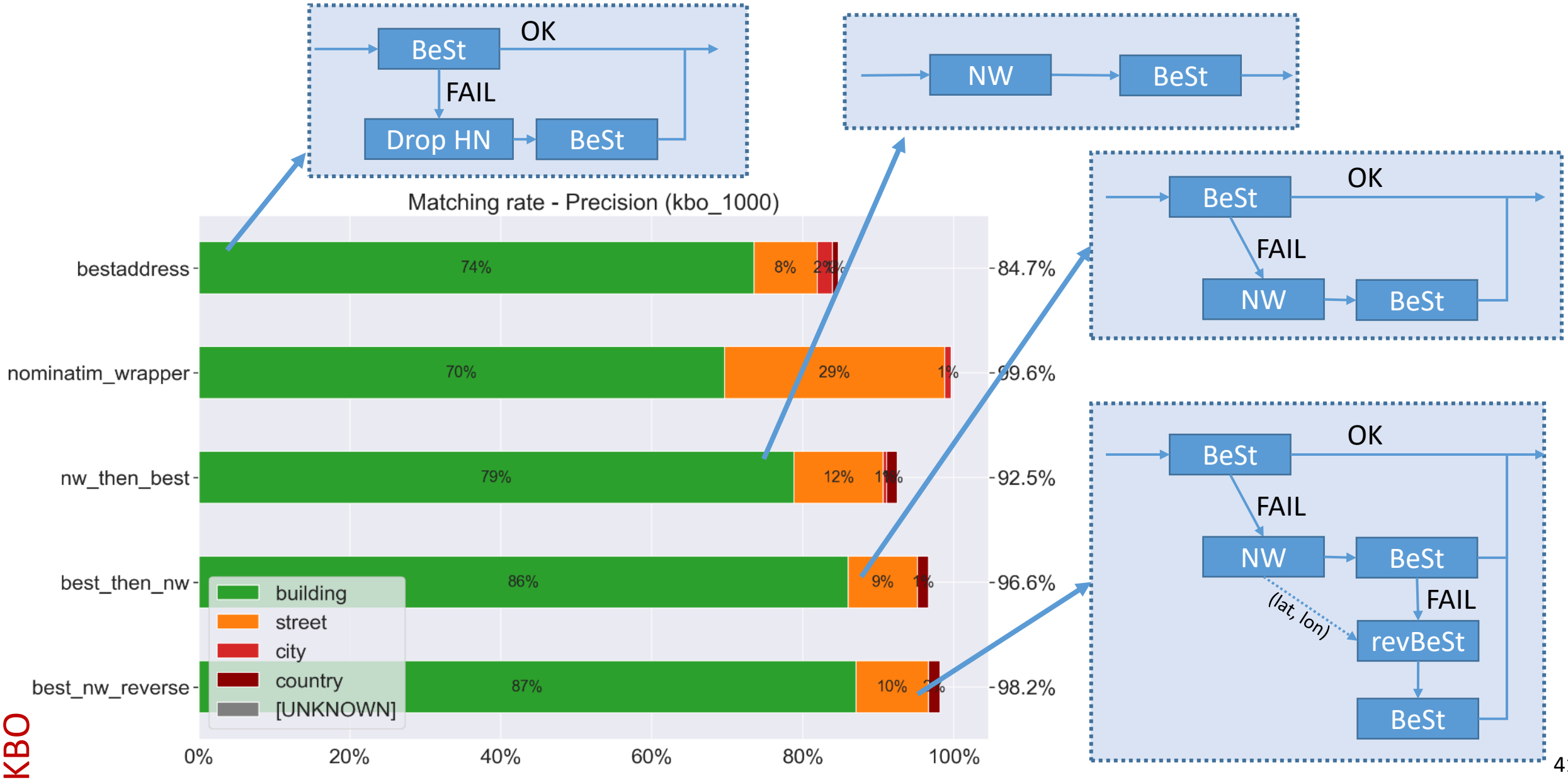
FOD Beleid & Ondersteuning
SPF Stratégie & Appui

- BeSt Addresses: BOSA project providing an authentic source for Belgian addresses
- Gathering regional authentic sources: UrbiS, Adressenregister (ex CRAB), ICAR
- Each address has a (stable) id which should be used in other applications
- Can be used either through file download (public) or via API (eGov)
- Many services available (list of streets within a municipality, list of addresses within street…), but no real geocoding (no fuzzyness) → no result if query does not fully fit BeSt data (streetname, housenumber)

Smals
ICT for society

https://bosa.belgium.be/fr/services/best-address-services

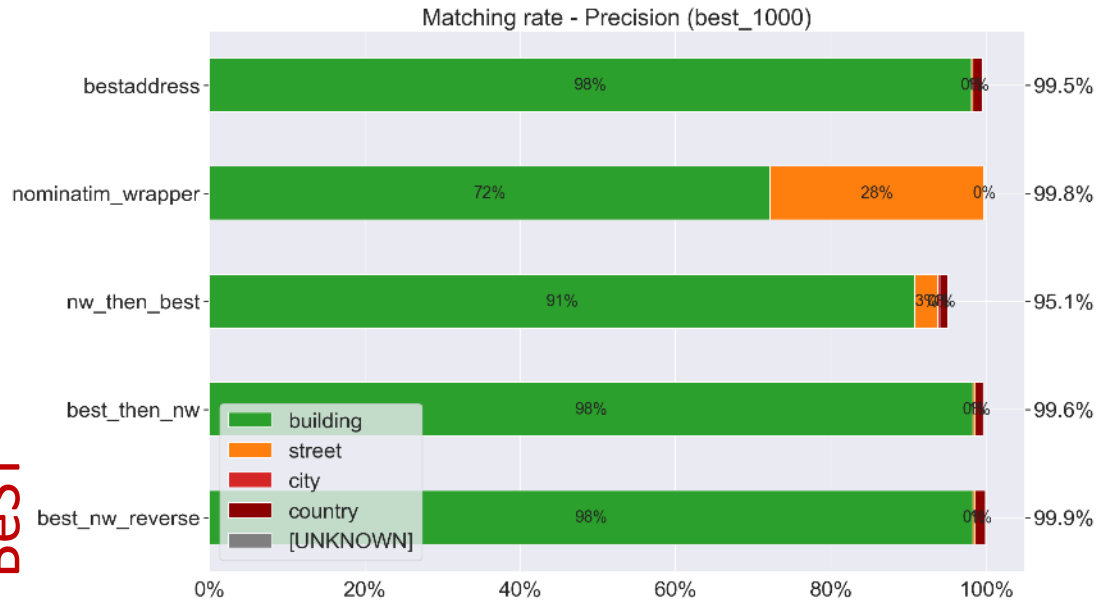
BestAddress and NominatimWrapper ?



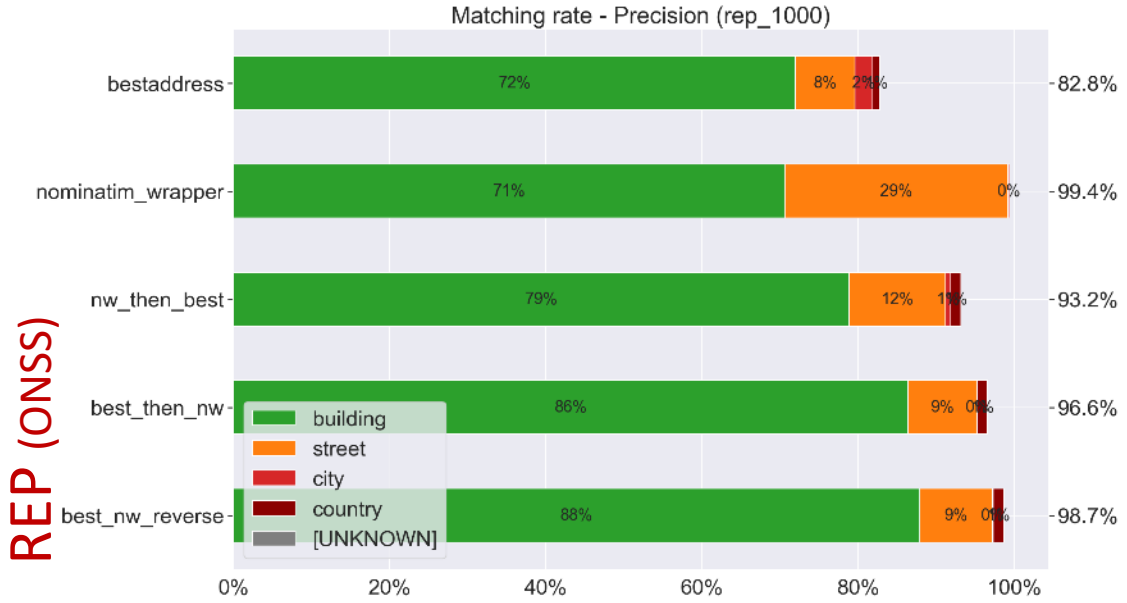
KBO

BestAddress and NominatimWrapper ?

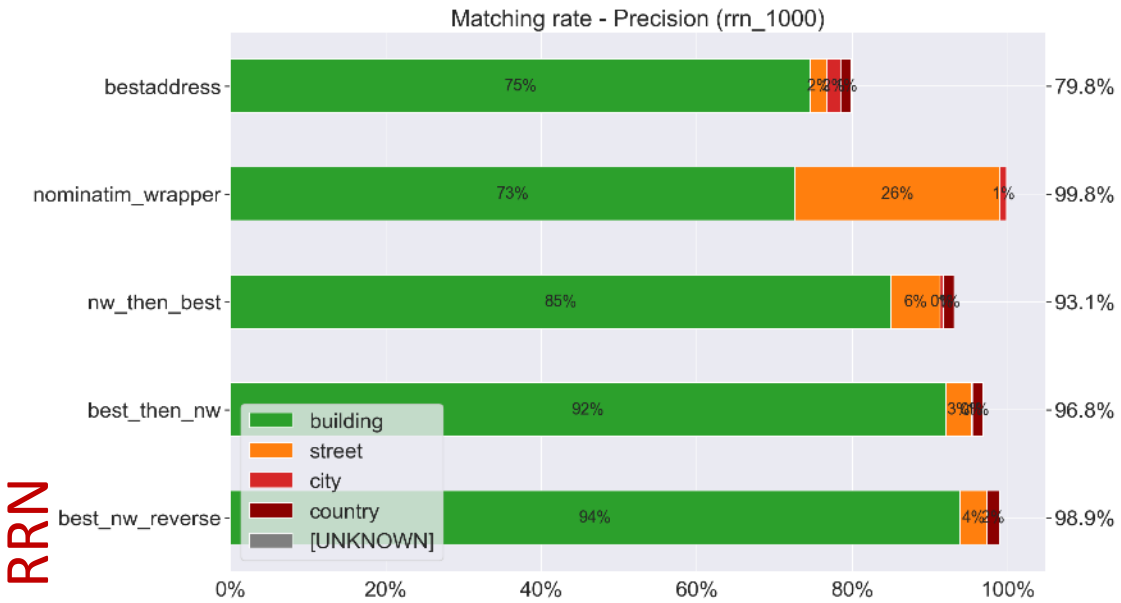
BeST



REP (ONSS)

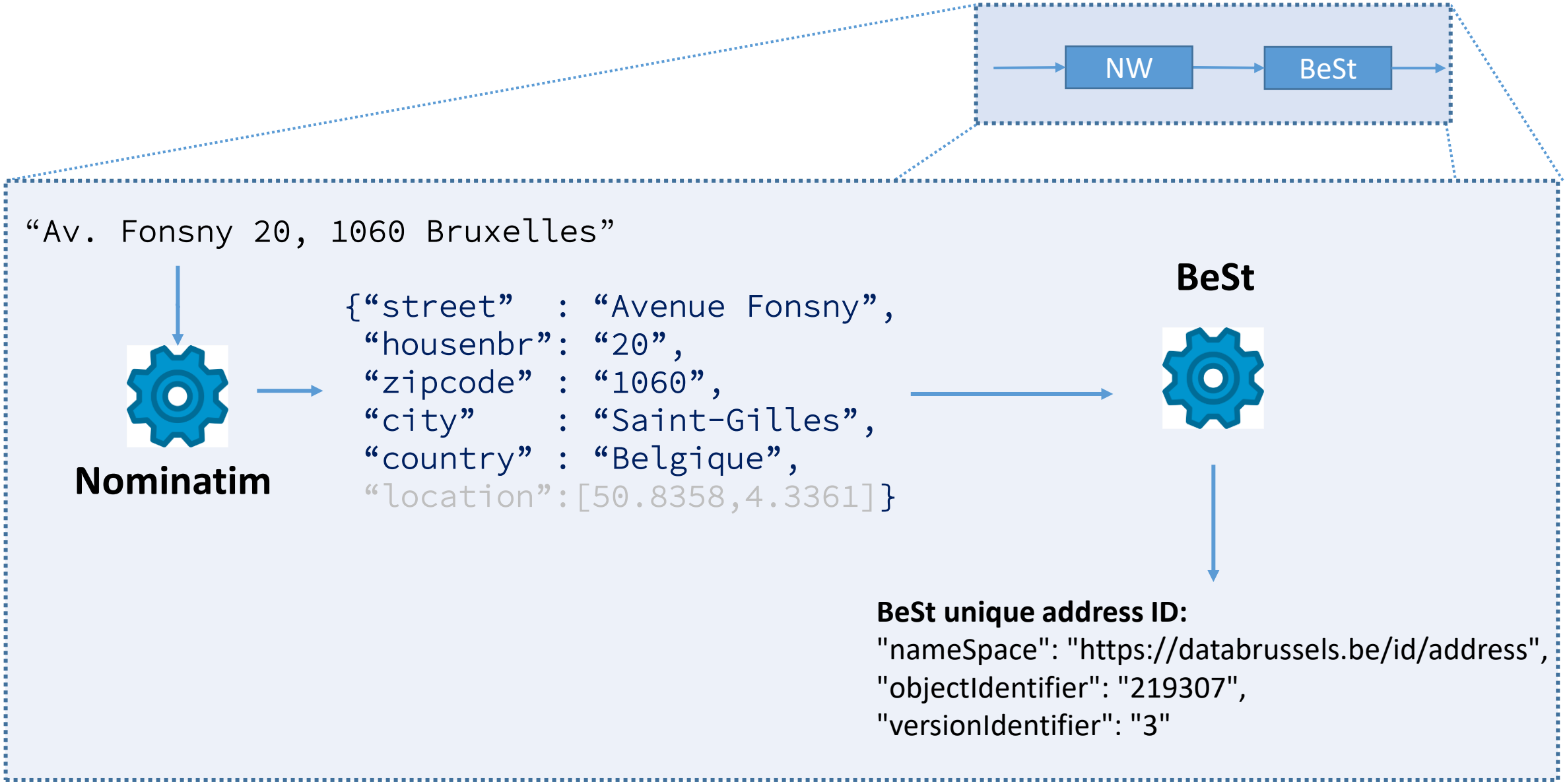


RRN



Smals
ICT for society

BestAddress and NominatimWrapper ?



BestAddress + NominatimWrapper

- Allow to add the « address cleansing » part to BeSt API
- This is not implemented in NominatimWrapper!
- It was just tested in a « lab » environment as a PoC
- With a script with a few dozen of Python lines

Smals
ICT for society

Agenda

- Concepts
- Evaluation
- Nominatim based solution
- BestAddress+Nominatim based solution
- **Pelias based solution**
  - **Pelias**
  - **PeliasPlus architecture**
- Conclusion

Map: cartesius.be (KBR)

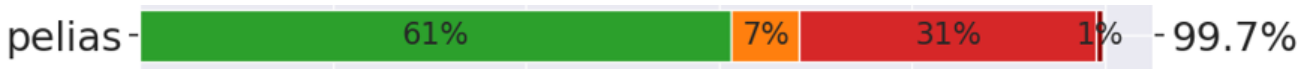
Smals
ICT for society

Pelias

- Open source: https://github.com/pelias/docker (projects > belgium)

- Data-source problems:
  - Based on OpenAddresses.io version of BestAddress CSV files (https://batch.openaddresses.io/), but due to some incompatibility, not updated since Feb 2021 (work in progress)
  - Pelias still uses an old OpenAddresses dataflow, deprecated mid-2021

- Robustness problem:



  - If structured version doesn't work, sometimes unstructured version does (and vice-versa)
  - If none works, some simple transformations often allow matching:
    - {"postalcode": 1160, "locality": "Auderghem"} → {"postalcode": 1160}
    - Cleansing text, removing parenthesis in street, non-digits in housenumber…

- Ongoing project (NGI , Bosa, Crisis Center, Smals) to address those problems

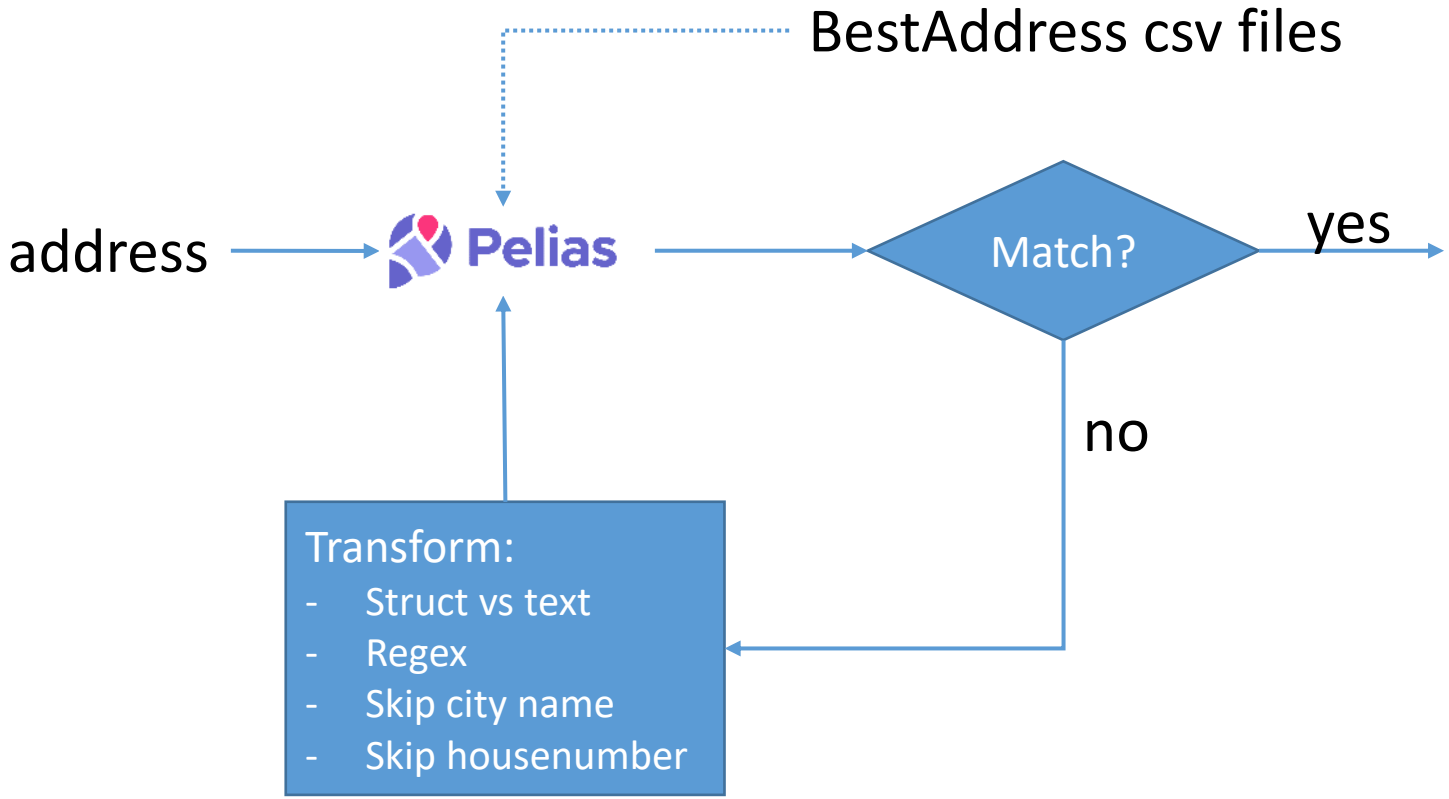
NGI Nationaal Geografisch Instituut — IGN Institut Géographique National

Centre de crise National

BOSA FOD Beleid & Ondersteuning SPF Stratégie & Appui

Smals ICT for society

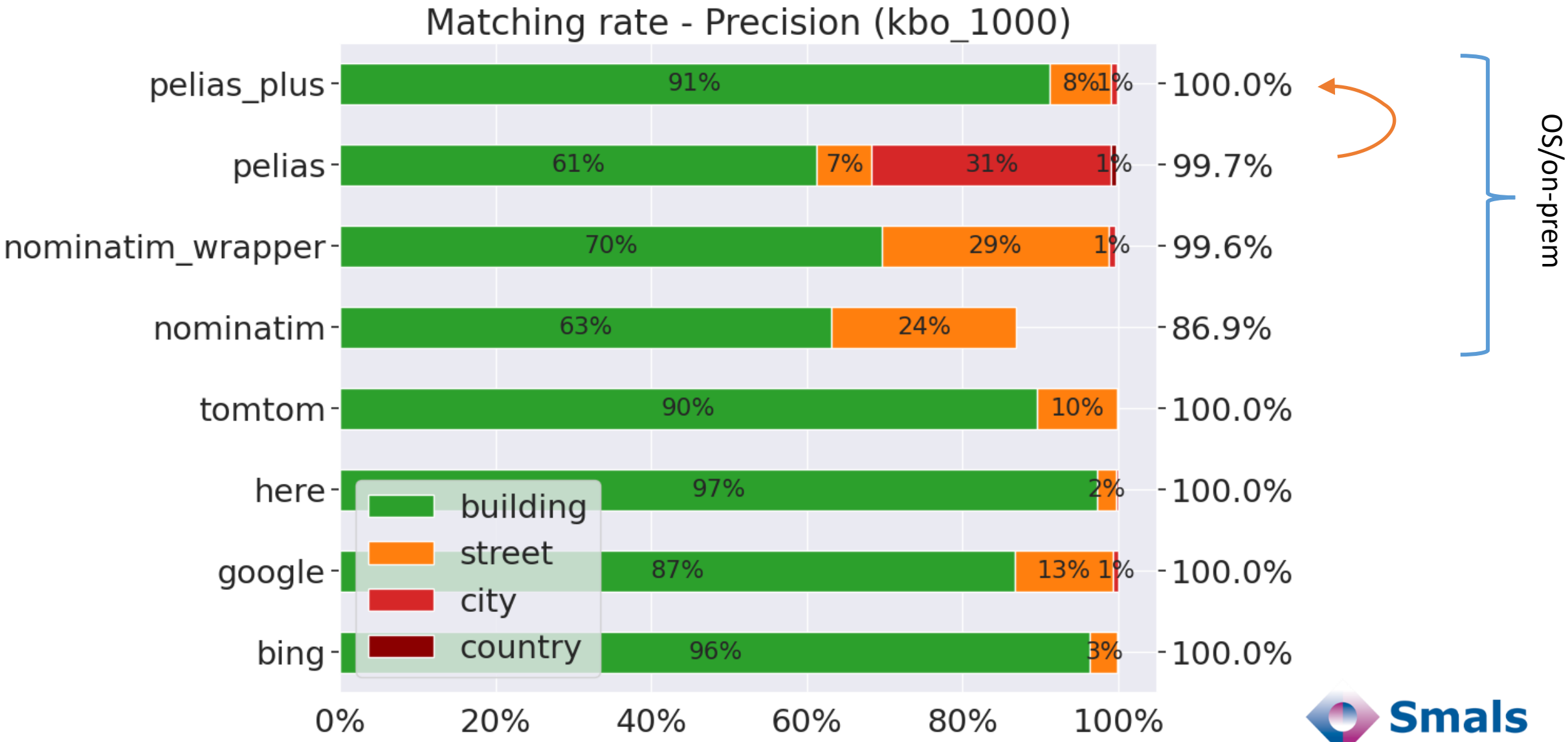
PeliasPlus



- Very early stage!
- Based on a new dataflow from BestAddress, by-passing OpenAddresses.io
- Same « transformer » logic as NominatimWrapper
- If housenumber not found:
  - Often close number is known
  - Example: “10” KO, but “10 A” or “8” and “12” OK → interpolation!
- Input data flow to be improved (e.g., municipality names + part of municipality)

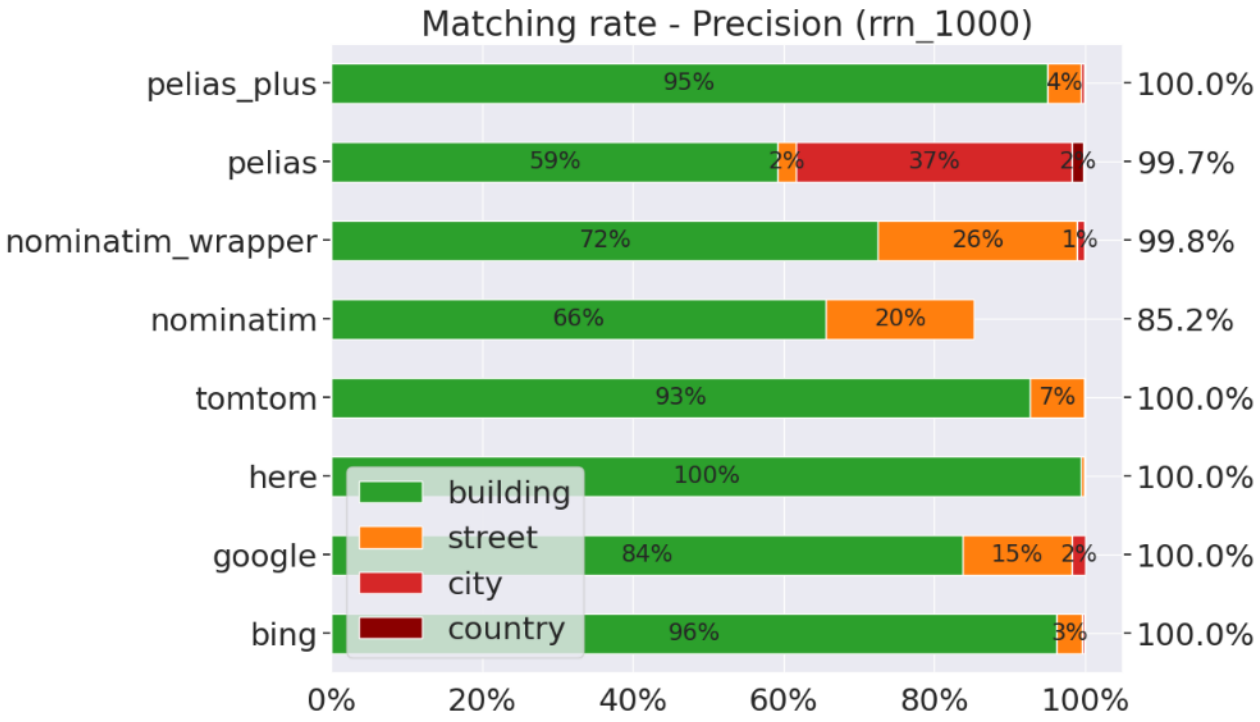
Smals ICT for society

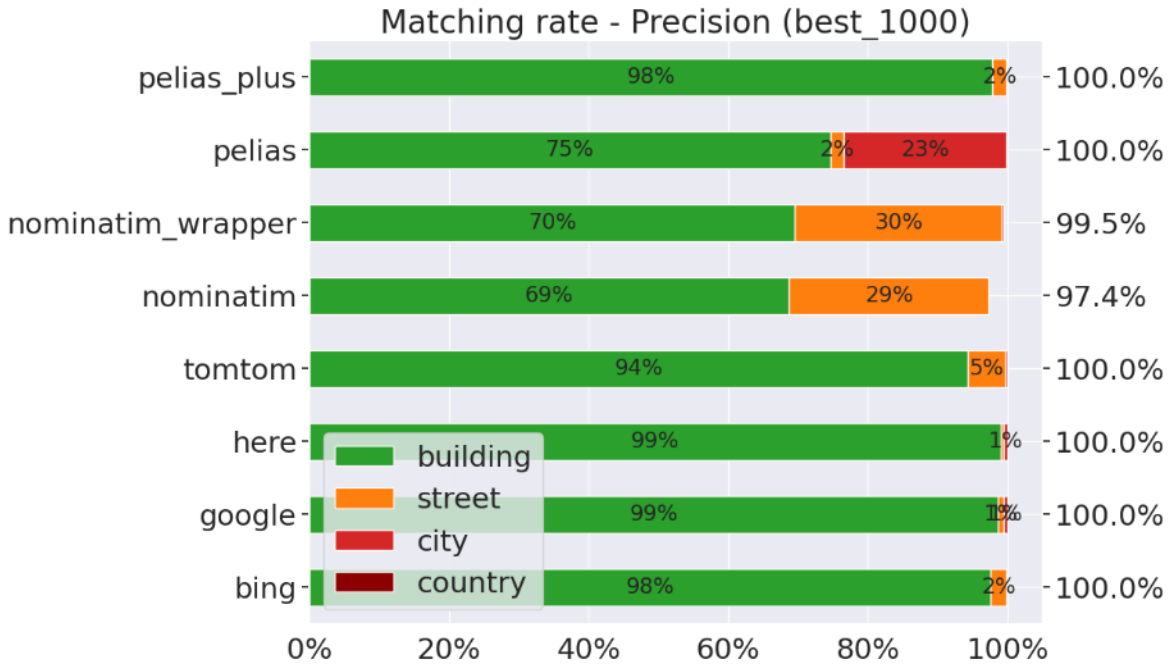
PeliasPlus: Matching rate

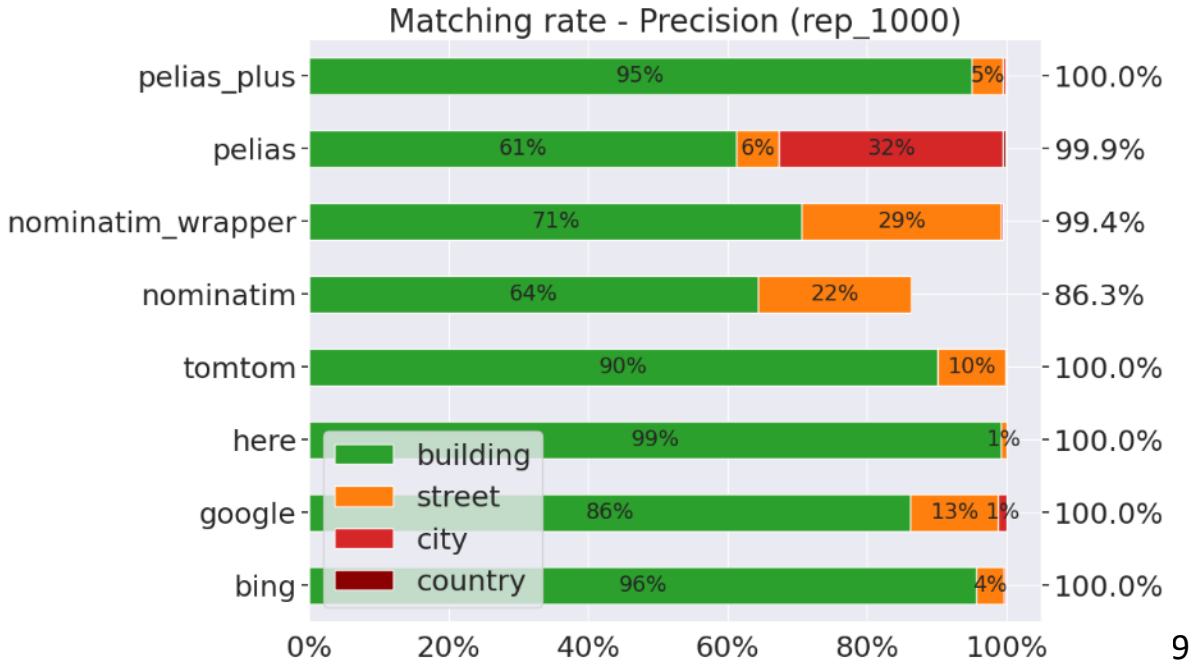


Experimental results!

PeliasPlus: Matching rate

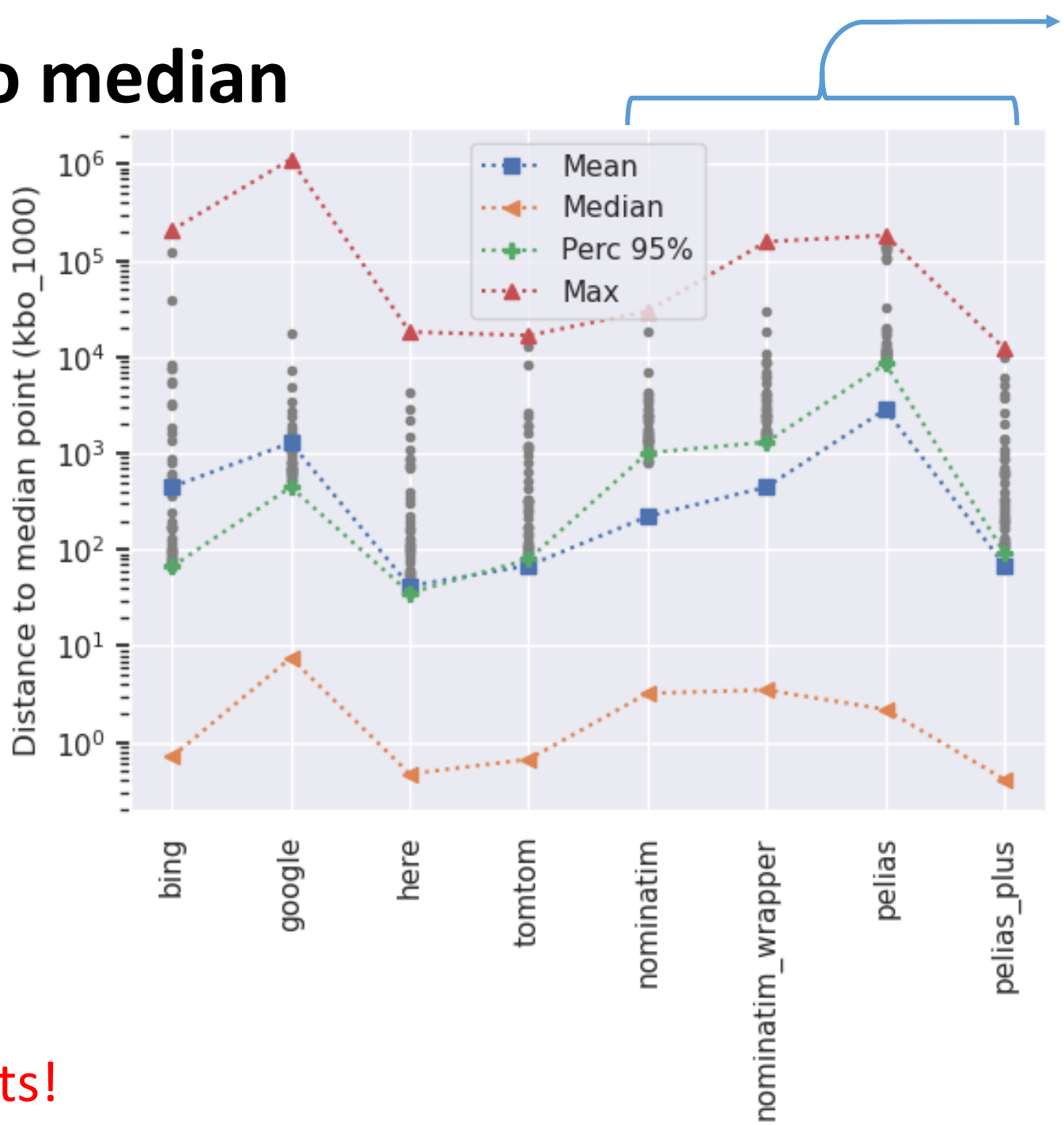






Experimental results!

Distance to median

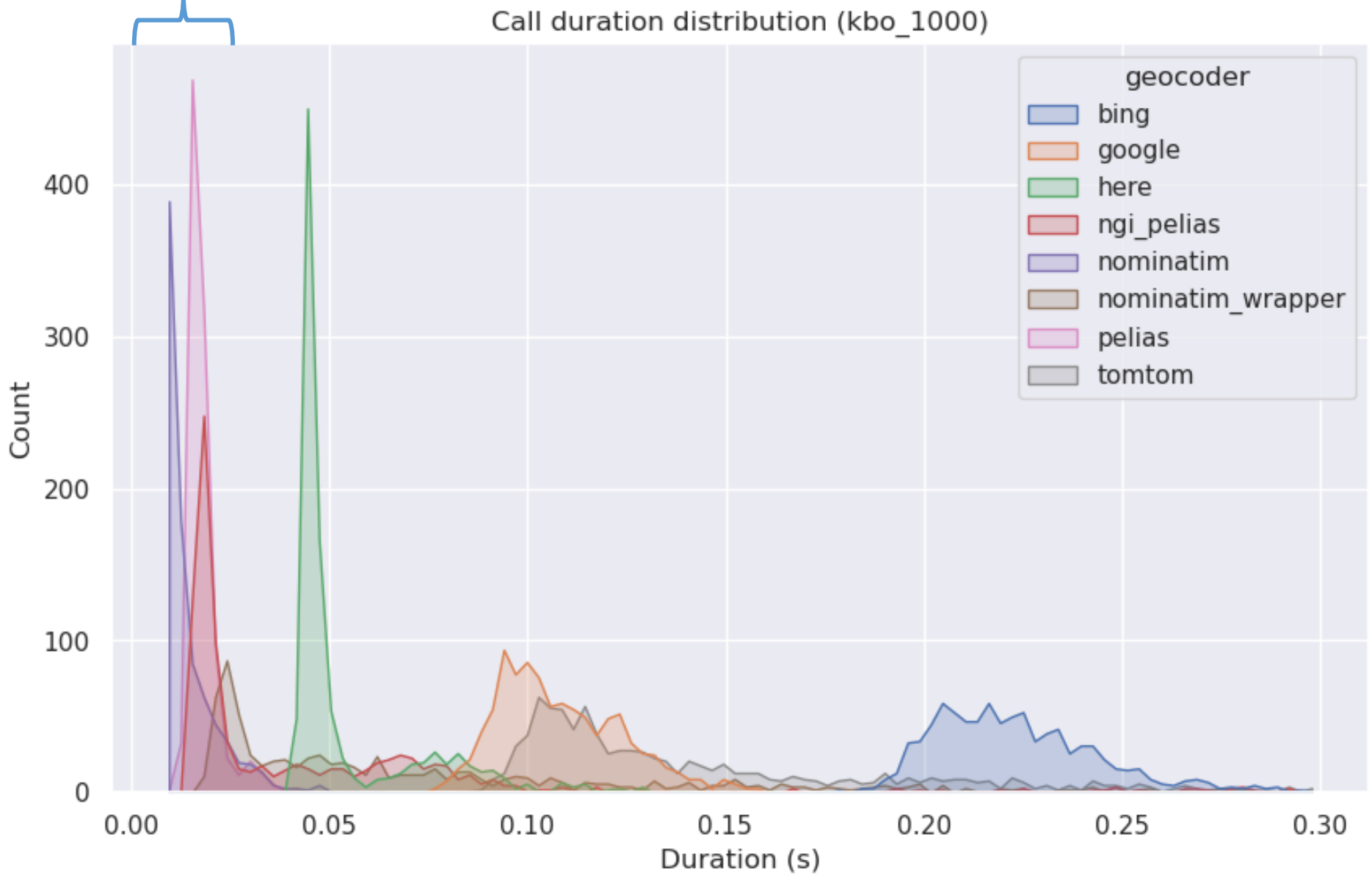


Experimental results!

Smals
ICT for society

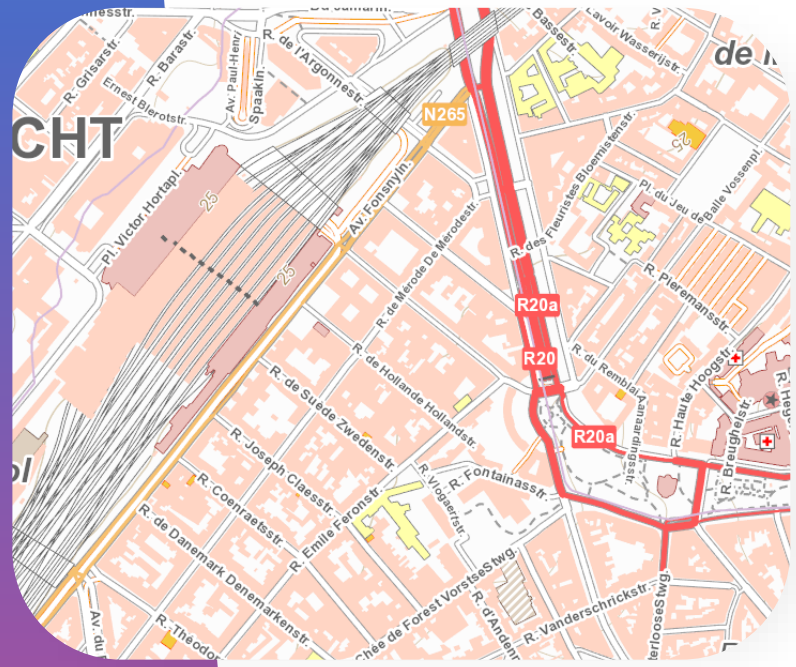
Call duration

OS/on-prem (→ same machine!)



Smals
ICT for society

Agenda



- Concepts
- Evaluation
- Nominatim based solution
- BestAddress+Nominatim based solution
- Pelias based solution
- **Conclusion**

Map: geo.be (IGN-NGI)

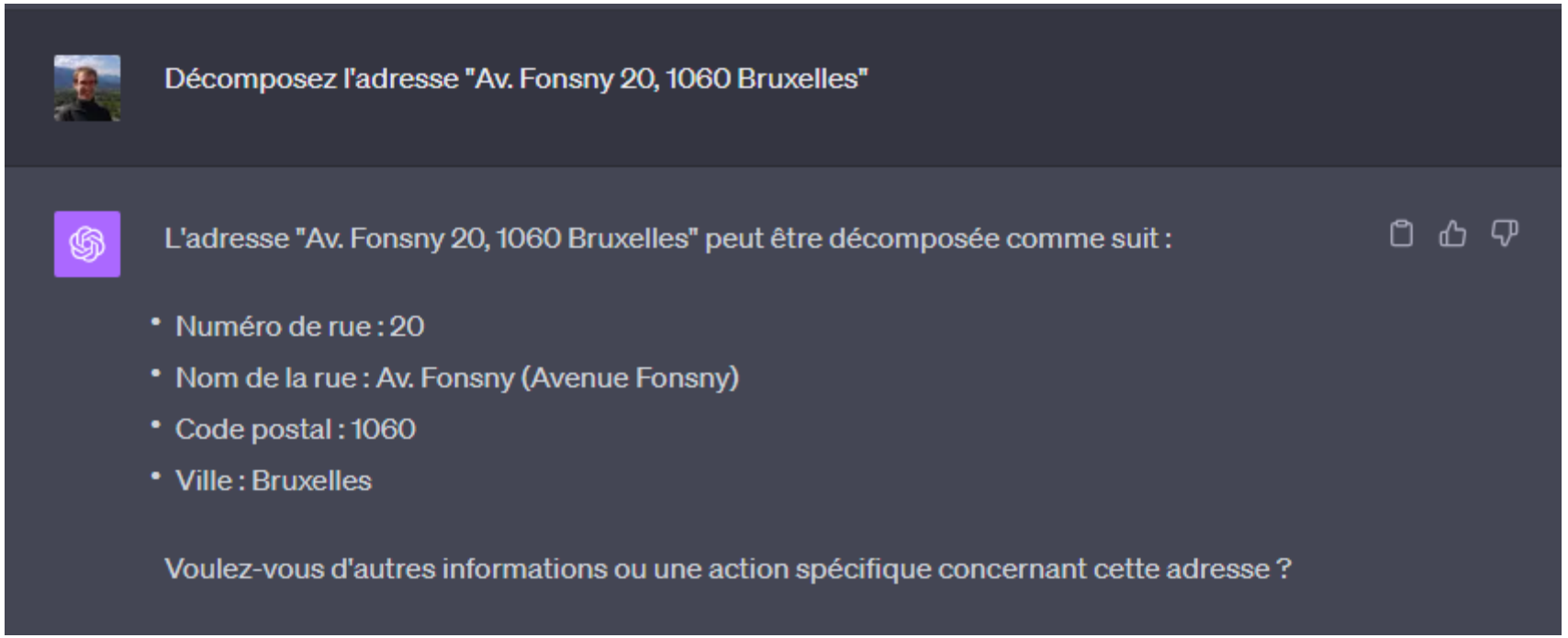
Smals
ICT for society

External partner

- To be also considered: external (large) company providing GIS solutions (ESRI, IBM…)
- Additional overhead to test/evaluate them
- Worthy for only geocoding?
- Perhaps useful if an SLA needs to be provided

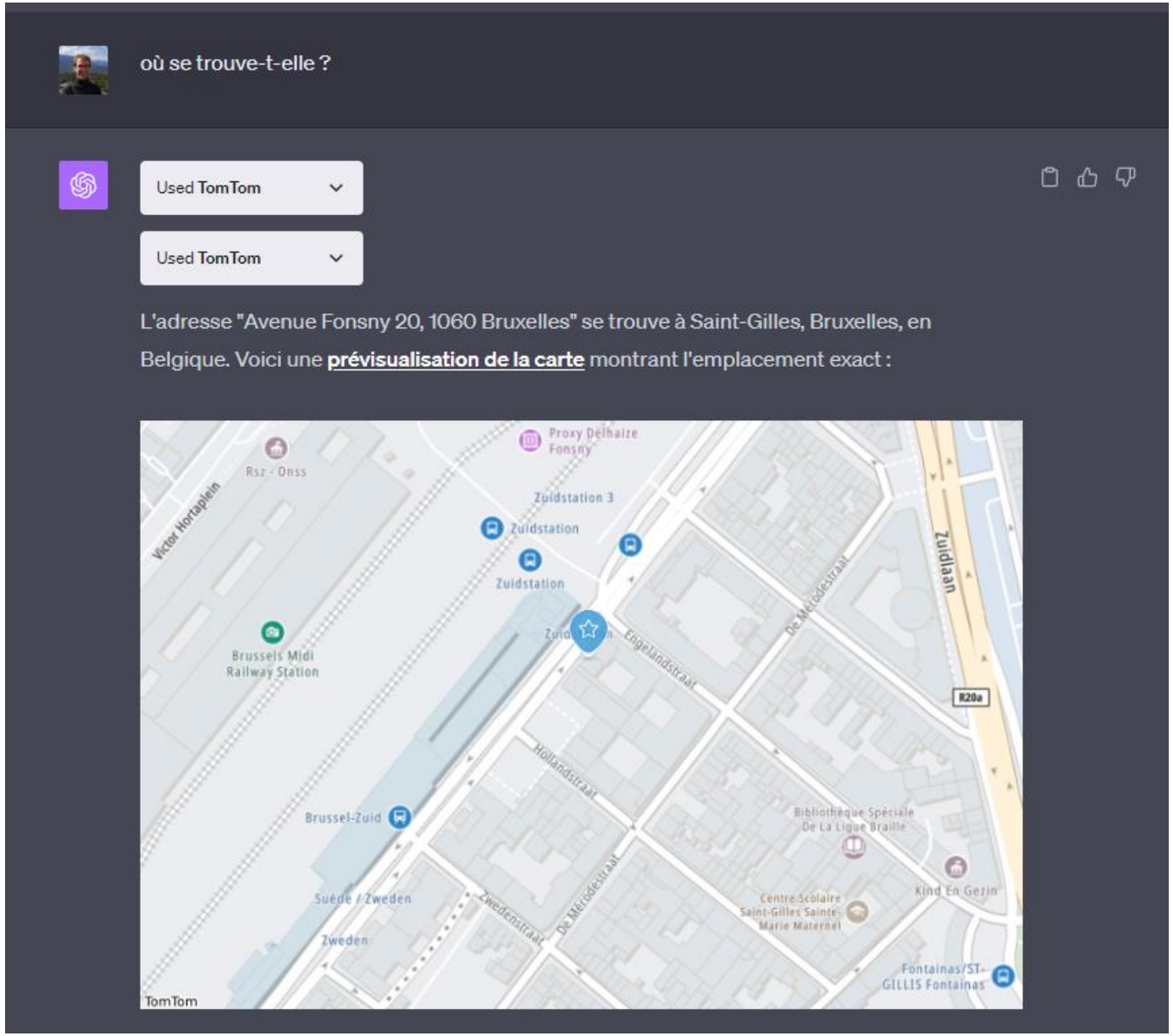
Smals
ICT for society

Chat GPT and geocoding



Smals
ICT for society

Chat GPT and geocoding (+ plugin TomTom)



Smals
ICT for society

Comparisons

	Comment	Pros	Cons
Cloud solutions	Variable quality!	Easy set-up Worldwide	GDPR limitations Requires internet access Costs for large volumes
Open source solutions	Not good enough for production apps	On premise Free usage	Requires infra & management
Adapted OS solutions	Based on OS solutions, but (almost) reach cloud solution quality	On premise Free usage	Requires infra, management & maintenance

Smals
ICT for society

Comparisons – best of

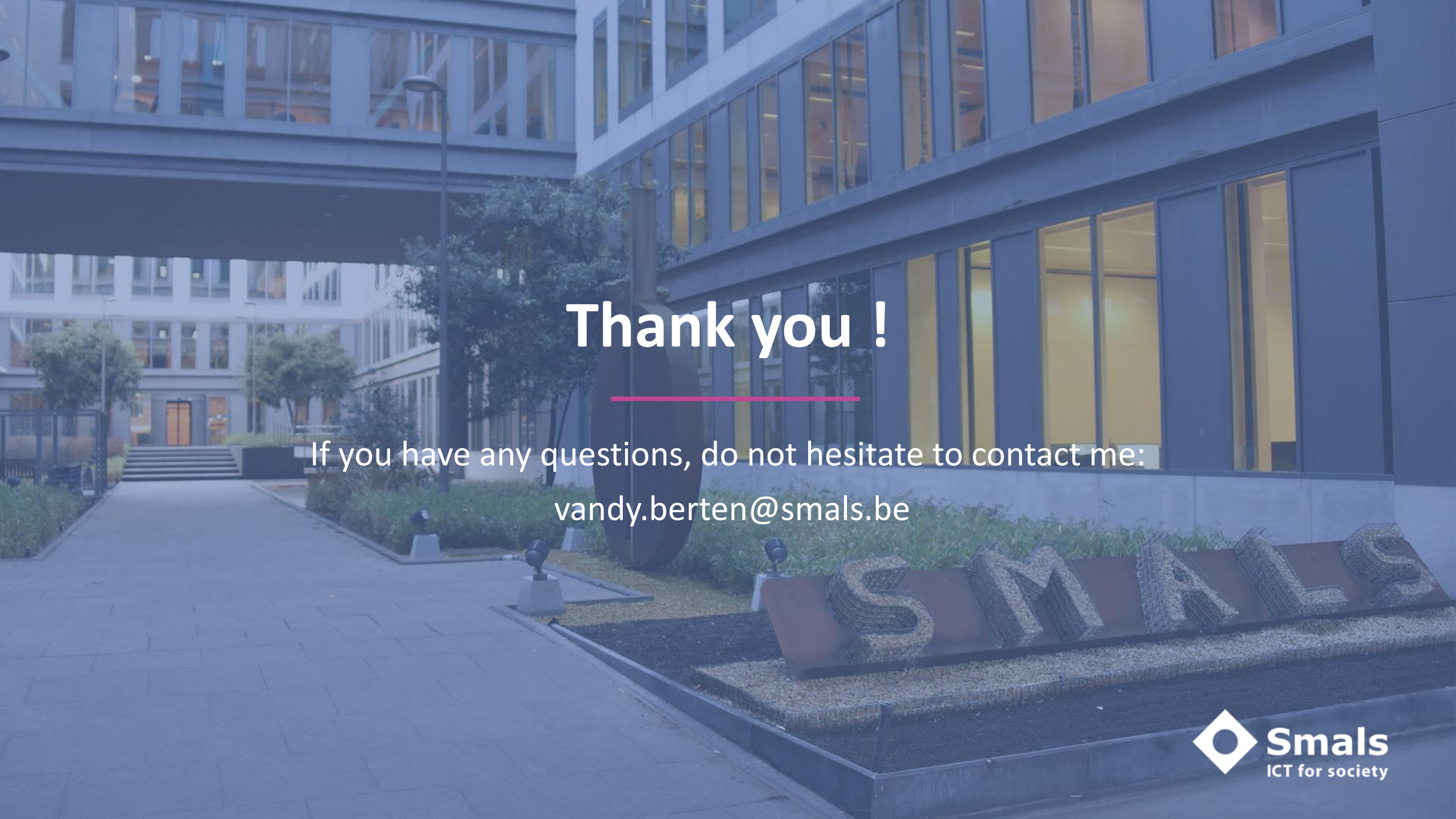
	Pros		Cons	
Here	Good quality Easy set-up Free up to 30 000 calls/month Worldwide		GDPR limitations Requires internet access Not for large volumes	
NominatimWrapper	On premise Free usage Very fast	OpenStreetMap data Works for “any” country	Requires infra & maintenance	Often street level
PeliasPlus		Good quality BestAddress id/data		Not mature yet Only Belgium

Smals
ICT for society

Conclusions

- Choosing a good geocoder is a complex task!
- Pick a (good) cloud solution if you can!
- Hybrid solution is possible (tool 1 for Belgium, tool 2 for World)
- Focus on address standardization or geolocation can lead to different tools
- Follow the NGI geocoder project evolution!

Smals
ICT for society



Thank you !

If you have any questions, do not hesitate to contact me:

vandy.berten@smals.be

Smals
ICT for society