





Suurandmed statistikas

Erki Saluveer
www.positium.com





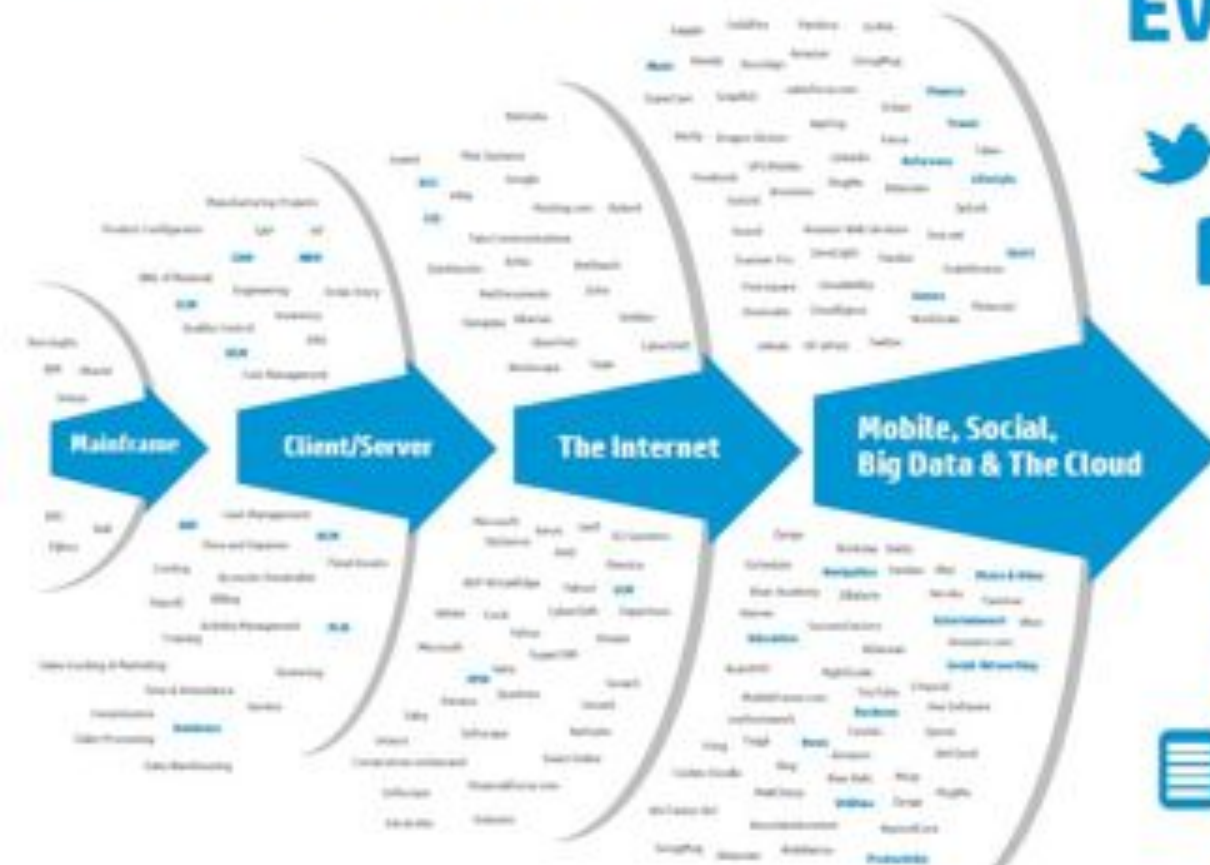
Suurandmed? BIG DATA?

Andmemahult suur (volume)



Kiire (velocity)

A new style of IT emerging



Every 60 seconds



98,000+ tweets



695,000 status updates



11million instant messages



698,445 Google searches



168 million+ emails sent



1,820TB of data created



217 new mobile web users

Struktureerimata (variety)



Scheveningen Memorandum „Big Data and Official Statistics“ DGINS

*Acknowledge that Big Data represent new opportunities and challenges for Official Statistics,
and therefore encourage the European Statistical System and its partners to effectively examine the potential of Big Data sources in that regard.*

EUROSTAT Task Force ‘Big Data and Official Statistics’



Feasibility study on the use of mobile positioning data for tourism statistics

EUROSTAT Feasibility Study on the Use of Mobile Positioning Data for Tourism Statistics

- Olemasolevate ametlike turismiandmete puudused:
 - Ajakohasus (tihti kompromiss ajakohasuse ja kvaliteedi vahel)
 - Asjakohasus (sõltuv olemasolevatest andmetest, mitte vajadustest)
 - Täpsus (valimid, geograafiliselt piiratud või täpsus varieerub)
 - Terviklikkus (sise ja välisturismi mõõdetakse erinevalt ja ei moodusta tervikut)
 - Koherentsus (erinevad meetodikad)
 - Kulu ja vaev (administratsioonilised kulud ja küsitluste keerukus)

Passiivne
mobiilpositsioneerimine –
telefonide asukohad
mobiilioperaatori logifailidest



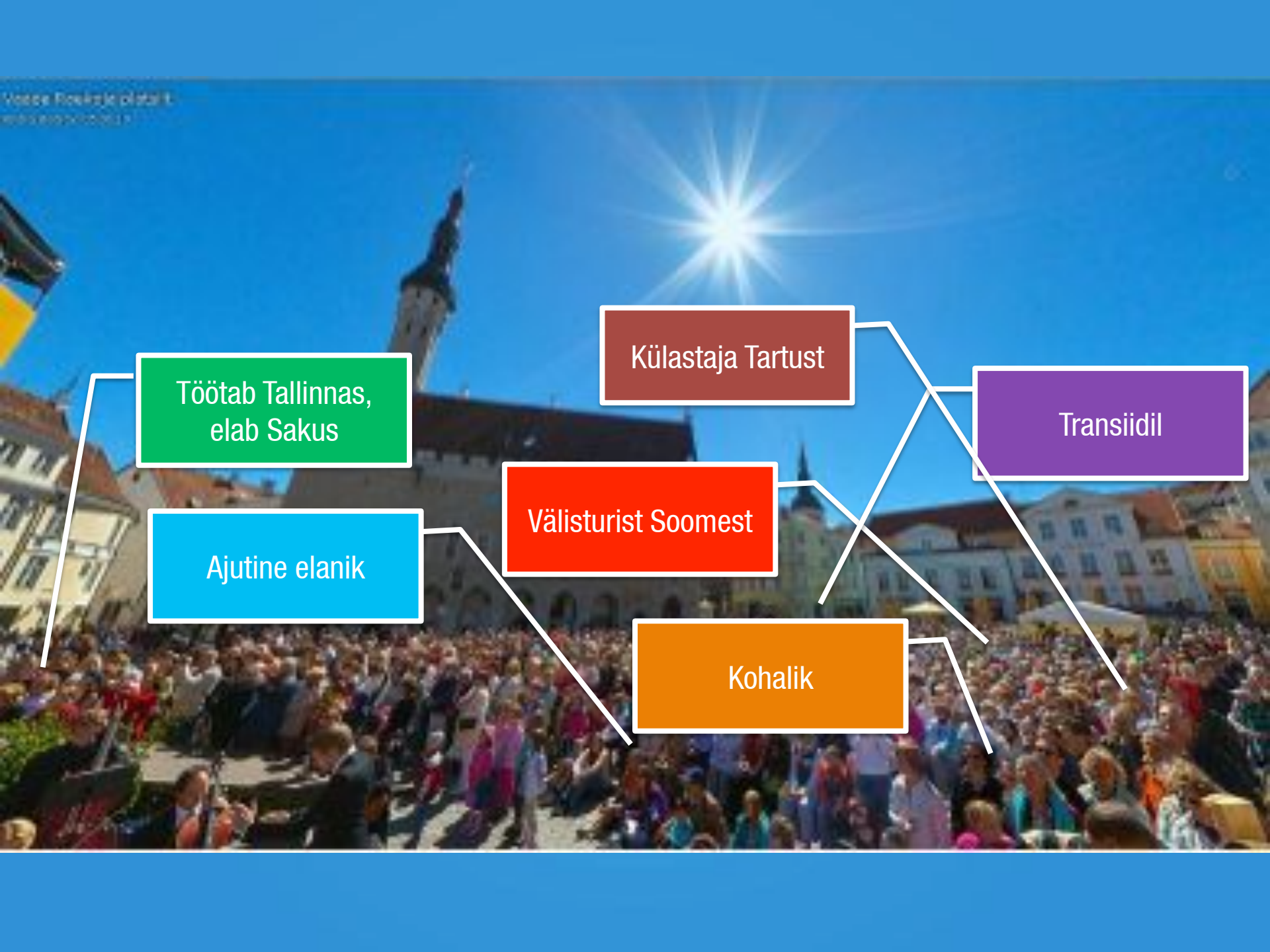
CDR näide



id	time	cgl
244051326004	2014-10-10 19:05:00	47203
234153291055	2014-10-12 11:20:06	44890
234153297005	2014-10-13 14:45:00	43948
244211635440	2014-10-04 14:02:36	559
250023198539	2014-10-13 19:59:41	49035
270013206233	2014-10-13 15:51:21	40734
250023201900	2014-10-07 20:50:46	590
250022229700	2014-10-13 13:11:24	20826
242012177204	2014-10-08 13:13:06	61037
244052103039	2014-10-01 11:01:02	44131
244213200159	2014-10-10 21:09:59	50378
244053264011	2014-10-06 16:33:05	246
244051300035	2014-10-13 12:05:22	40950
270013295051	2014-10-13 21:49:22	45464
234153297176	2014-10-13 00:12:30	559
244051334443	2014-10-01 10:57:27	227
244212129277	2014-10-04 15:00:10	1803
244051257212	2014-10-14 11:36:20	43951
244052062510	2014-10-04 22:00:04	43908
24412606294	2014-10-05 23:55:41	461
202011431701	2014-10-07 13:10:31	44204
270011096501	2014-10-05 01:46:50	60974
234153291137	2014-10-11 01:21:22	50200
244211121763	2014-10-03 15:34:06	46326
226011029311	2014-10-13 10:21:35	15509
244211049600	2014-10-15 09:04:45	43951
244053274061	2014-10-08 09:22:47	33243
244051467065	2014-10-14 16:49:38	45663
244052694522	2014-10-03 12:25:10	43877
250023019927	2014-10-05 00:06:03	60792
250022010036	2014-10-06 10:33:46	39
230013297960	2014-10-13 14:29:19	45093
24405431258	2014-10-06 13:49:33	573
244212114066	2014-10-04 20:50:25	40907
244211325479	2014-10-02 19:21:13	30806
244121320901	2014-10-03 10:47:00	47157
244051247205	2014-10-05 09:25:03	234
244052391215	2014-10-13 16:15:27	60390
244212577245	2014-10-08 21:07:32	40907
244052106197	2014-10-04 01:36:20	1047
230012033323	2014-10-09 17:49:41	43610
244213316019	2014-10-15 12:46:59	43877
242011296004	2014-10-01 16:50:05	44431
244213293012	2014-10-11 20:37:44	227

Mustrid





Töötab Tallinnas,
elab Sakus

Ajutine elanik

Külastaja Tartust

Välisturist Soomest

Kohalik

Transiidil

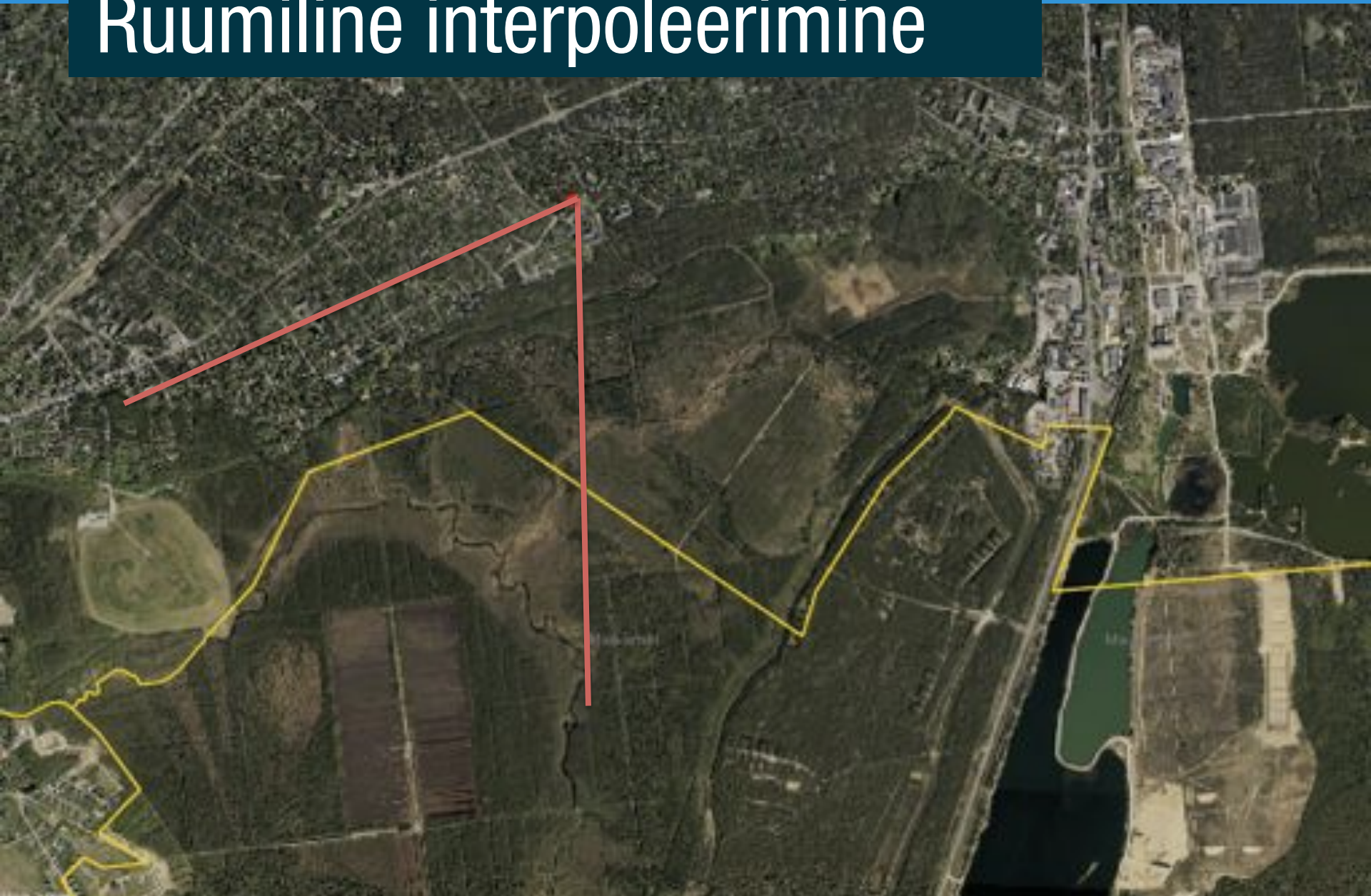
Müra



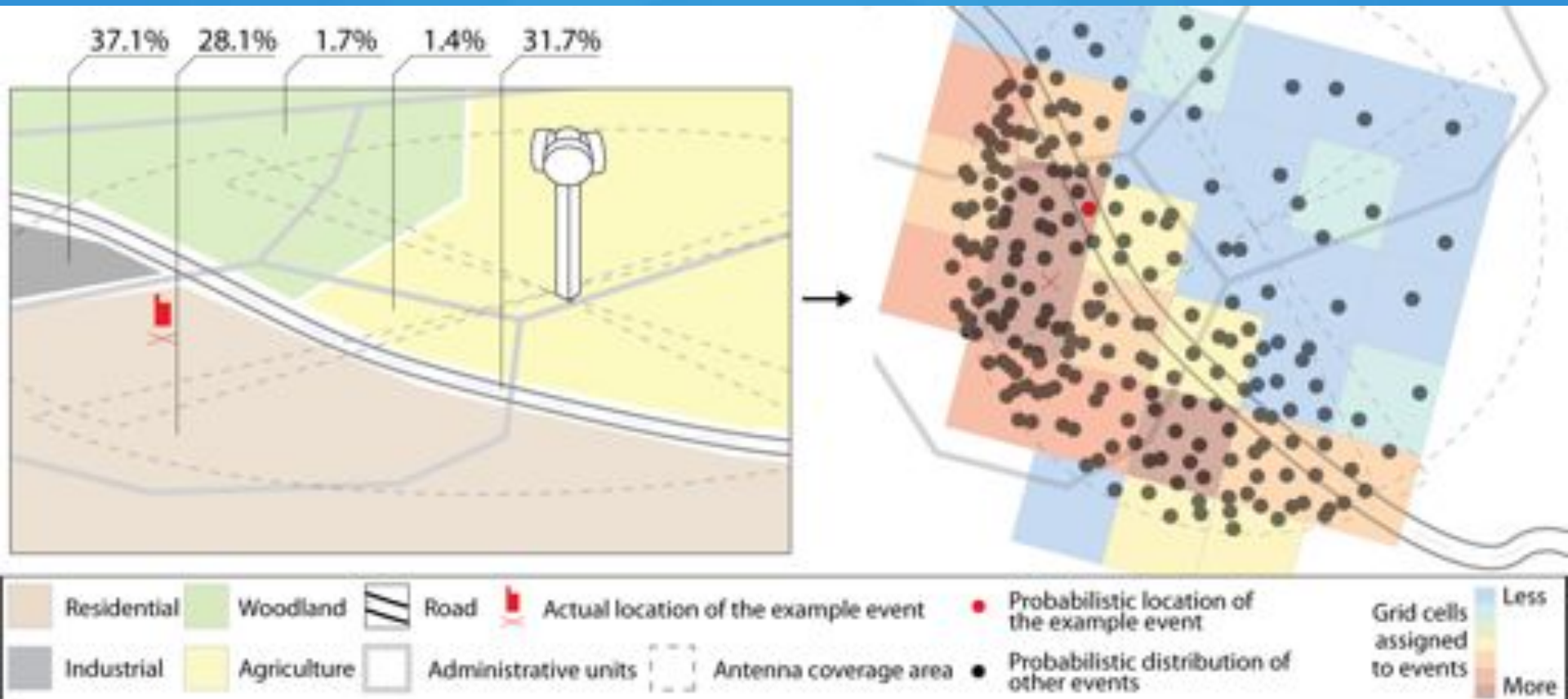
Müra ja andmevead

- Mobiiltelefoni ei kasutata (andmeaugud)
- Mobiiltelefoni kasutatakse liiga palju
- Kasutatakse kahte telefoni (twin-phones)
- Piiri ääres kasutatakse telefoni teisest riigist või merel laevalt
- Koduvalve seadmed jt M2M seadmed
- Puuduvad andmed
- Valed koordinaadid või muud parameetrid
- Operaatori andmevead või muutuvad süsteemid
- Jne

Ruumiline interpoleerimine



Probabilistic geographic distribution



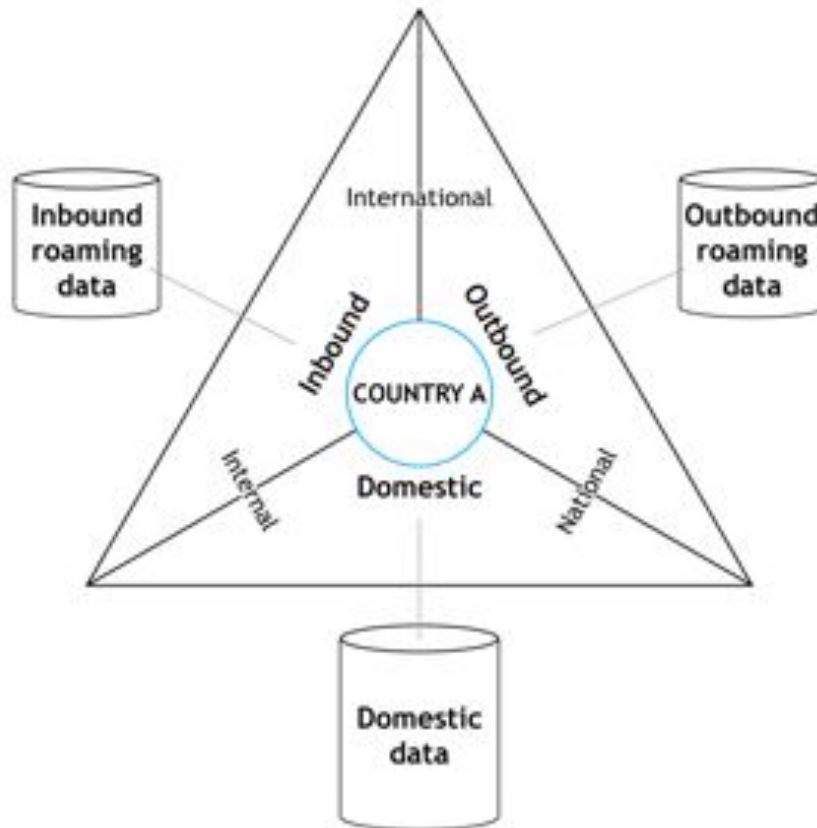
Positium Data Mediator

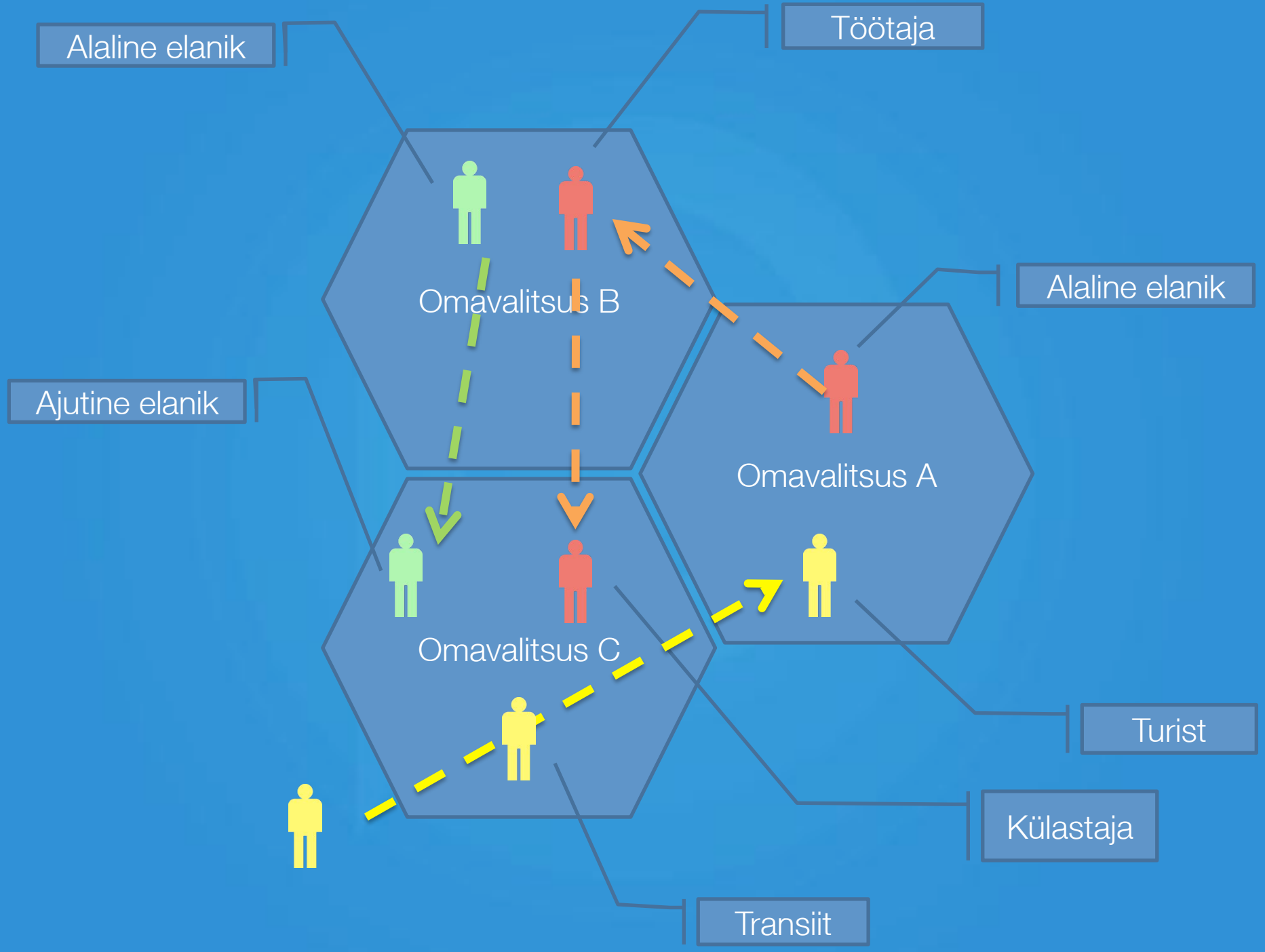


* Specific sources of data depend on operator's technical specs.

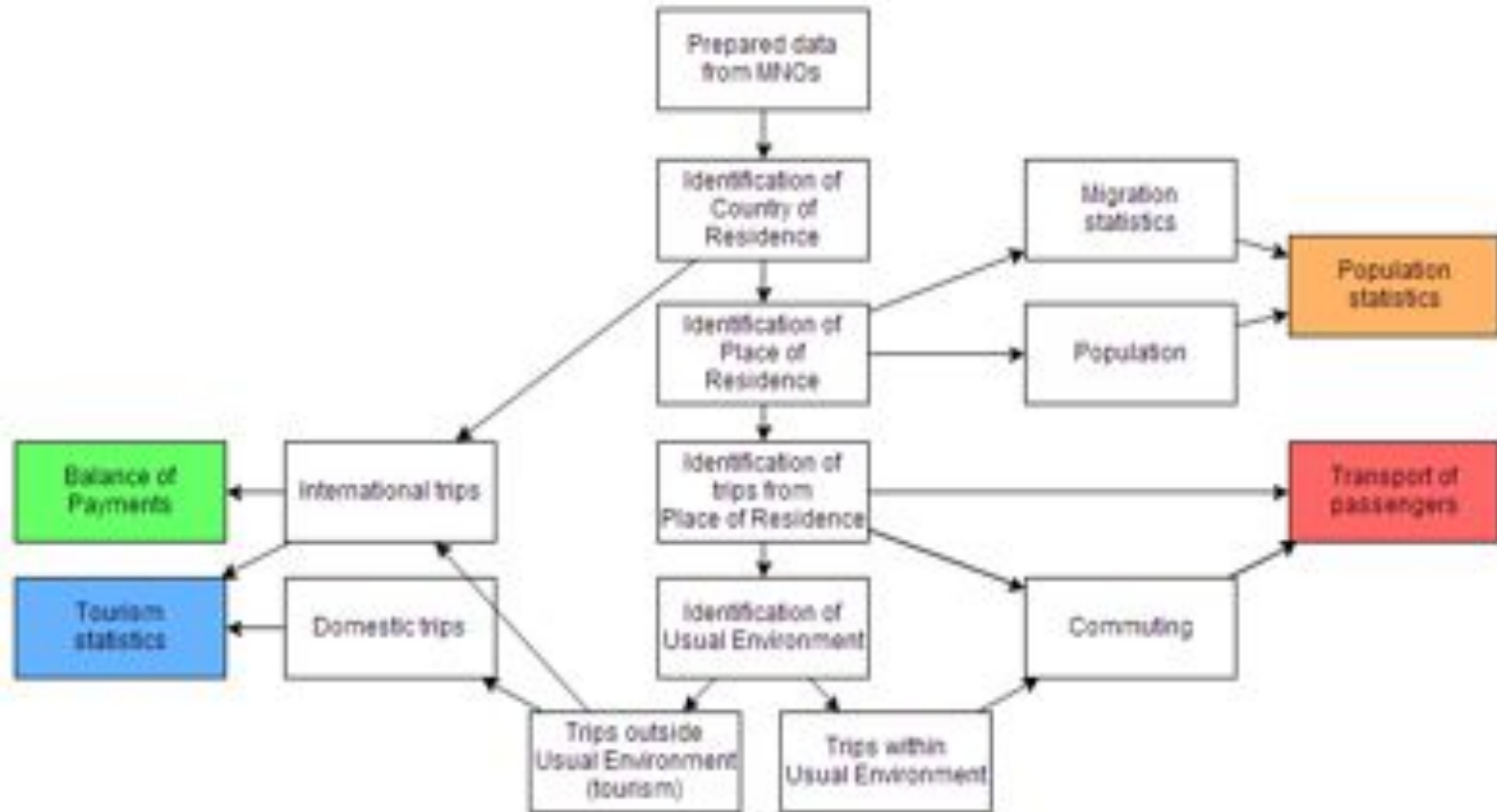
Positium © 2011

Turismistatistika allikad





Ametlik statistika mobpos andmetel



Turismistatistika indikaatorid

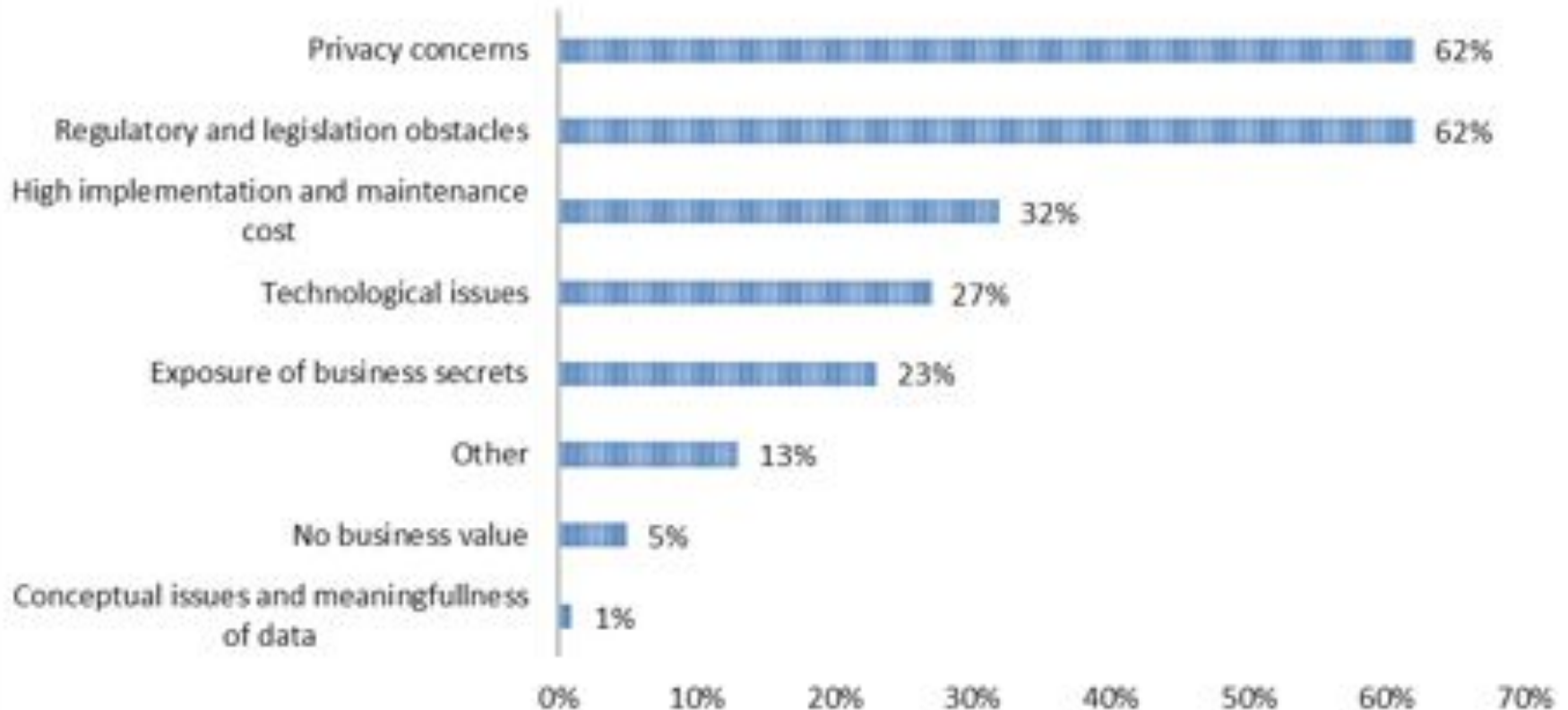
Klassifikatsioon:

- Elukohariik (COR), elukoht (POR)
- Aja agregaat (päev, nädal, kuu)
- Ruumi agregaat (grid, ov, aü, mk)
- Külastuse/Viibimise kestus (pikkus, ühepäeva/ööbimisega)
- Primaarne sihtkoht, muud sihtkohad, transiit
- Liikumismustrid
- Korduvkülastused

Kasutatavate indikaatorite näited:

- Külastuste arv
- Viibimiste arv
- Ööde arv
- Päevade arv
- Külastuse kestus
- Unikaalsete külastajate arv

Miks seda veel ei kasutata?



MobPos tugevused ja nõrkused

- Terviklikkus
- Kattvus
- Regionaalsed ja päritolu jaotused
- Kvantitatiivsed kriteeriumid definitsioonidele
- Paranenud ajakohasus
- Statistika genereerimise automatiseerimine
- Kuluefektiivsem
- Võimalus luua ühe metoodikaga üle-Euroopaline turismistatistika

- Juurdepääs andmetele
- Puudub informatsioon reisi eesmärgi, kulude ja transpordivahendite kohta
- Definitsioone ei ole võimalik täita 100% olemasolevatega (ühepäeva/mitmepäeva turist)
- Telefoni kasutamise mustrid ja sellest tulenevad üle ja alaesindatused
- Referentsandmed usaldusväärsuse hindamiseks puuduvad



EESTI PANK

Eesti Panga näide



4x

kiirem

12x

rohkem
riikide
eristust

100%

vähem turistide
tülitamist



Aitäh!

Erki Saluveer

<http://mobfs.positium.ee/>