



Open Data analysis to retrieve sensitive information regarding national-centric critical infrastructures

Renato Fontana

Research Project 2

Supervised by:
Ralph Dolmans,
Benno Overeinder



Presentation flow

- Introduction
- Known Studies
- Project Scope
- Visualization of Open Data
 - Demo
- Conclusions & Future Work
- Questions

Introduction

- Open Data initiative
 - Democratization of information;
 - Sharing and collaboration;
 - Application Development;
 - Accountability;
 - Freedom of Information Act;
 - Empowers users;
- Information disclosure;
- Sensitive content;



Source: OmgevingsAlert

“The FOIA establishes for any person—corporate or individual, regardless of nationality—presumptive access to existing, unpublished agency records on any topic.” *

* Critical Infrastructure Information Disclosure and Homeland Security, DTIC 2002.

Introduction

- Critical Infrastructures
 - What is critical ?
 - “Essential services that contribute to the stability and security of a Country” – Rinaldi 2001
 - Nation dependent
- CI Protection
 - National Security information
 - Information classification;
 - Sensitive but unclassified
 - Openness vs Secrecy

Research question

- Can users make use of Open Data databases to retrieve country sensitive information?
- How can this information be visually represented in a way that allows the easy identification of critical and strategically important areas on detailed level?

Known studies

- No studies in the visual representation of CI areas based on open data.
- “The risk of public data availability on critical infrastructure protection”. *Abbas, 2006.*
 - Compilation of SBU information;
- “Empirical findings on critical infrastructure dependencies in Europe”. *Luijff, 2009.*
 - Analysis on public reports of CI disruption events to recognize cascading effects.
- “Mapping the Dutch Critical IP Infrastructure”. *Alizadeh and Oprea, 2013.*
 - Dutch Critical Infrastructures organizations rely on foreign communication providers – based on public knowledge.

Project Scope

- Limited to Netherlands;
 - Existing repositories in National and City levels;
 - Country classification of CI;
- Limited dataset analysis;
 - Most relevant data reflecting CI;
 - Telecommunication & ICT;
 - Energy Resources;
- Visual representation;
 - Pattern recognition;

Open Data - Netherlands

- **Current view**
 - Netherlands 6th in OKFN rank.
- **Multiple reliable repositories;**
 - Government, private sector and user maintained;
 - data.overheid.nl;
 - maps.amsterdam.nl;
 - alliander.nl;
 - data.amsterdamopendata.nl;
 - Other public sources;
- **Machine readable;**
 - .csv, .xls, .json, live APIs;

Dutch Critical Infrastructures

Dutch government on the CIP identify the following as critical infrastructures



Energy



Health



Legal Order



ICT



Financial



Transportation



Water supply



Public Order
and Safety



Chemical,
Nuclear
Industries



Food Chain



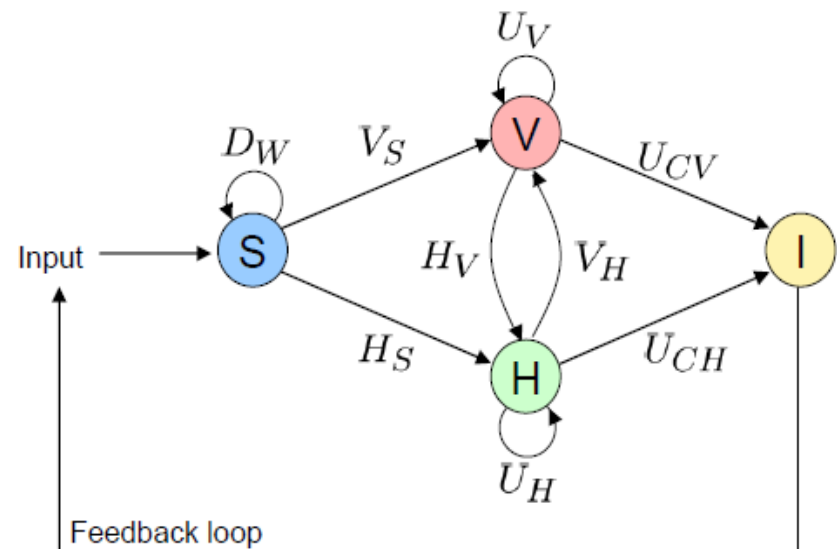
Public
Administration

Source: [government.nl/issues/crisis-national-security-and-terrorism](https://www.government.nl/issues/crisis-national-security-and-terrorism)

Approach

- Create hypothesis and visualizations;
 - Relevant CI Datasets;
 - Experimentation;
 - Gain insights;

- Feedback loop process. *Keim, 2008*



Visualization of Open Data

- Data Acquisition
 - Web crawlers, visual inspection, search engines, manual download of contents;
- Data Sanitation
 - Removal of arbitrary and blank entries;
 - Project scope;
- Layer creation
 - Visual representation of datasets to CI categories;
- Quickly identification of patterns

Data Analysis

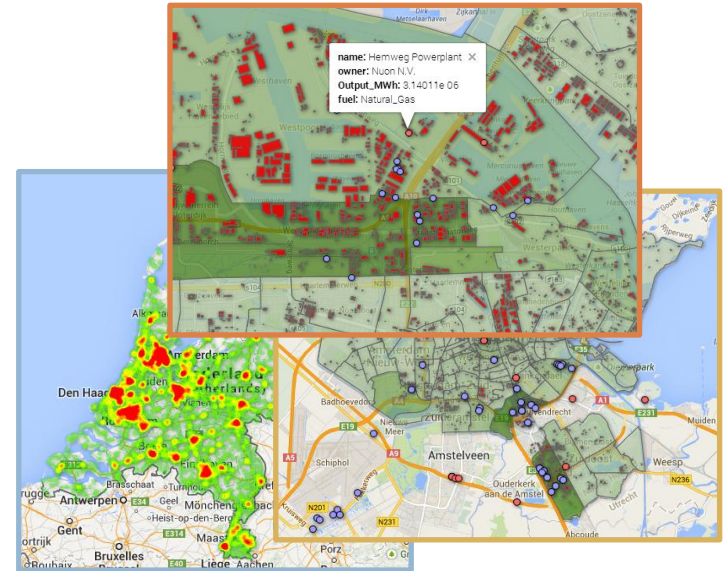
- Initial data inspection;
 - Common localization identifier among datasets
 - PostCode, Coords, LAT, LONG, City, *locatie*, *buurt*, *wijk*, *gem*
- Map representation;
 - Shape Files (.shp) parse;
 - Centraal Bureau voor de Statistiek – cbs.nl
 - Convert to .kml
 - Google Maps API;
 - Geocoding, Fusion Tables



Demonstration

Findings

- Patterns on CI;
 - Identified cities;
 - Amsterdam, Den Haag, Rotterdam;
 - Data Centers and Power Plants
 - NL-IX peering DCs, Hemweg PowerPlant;
 - High resource demanding areas
 - Westpoort, Sloterdijk, De Omval, Bullewijk
 - Details on electric consumption;
 - Business offices and companies
 - Interest in hiding data



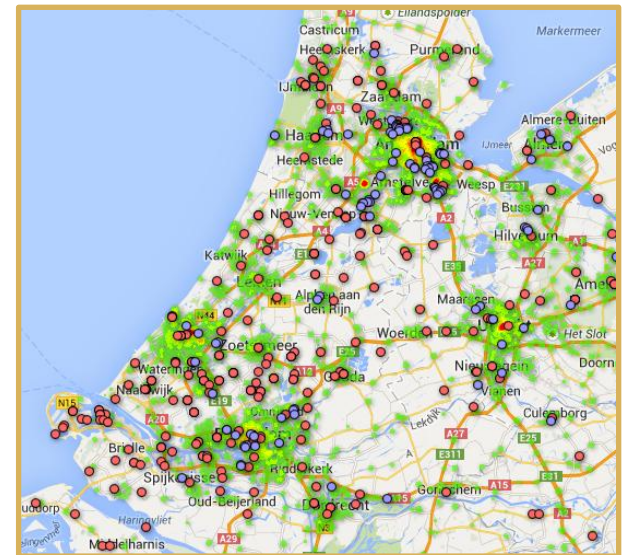
AMSTERDAM,1101A	Afgeschermd
AMSTERDAM,1101B	Afgeschermd
AMSTERDAM,1101C	22.961.527

Conclusions

- **Compilation of public information;**
 - Not trivial;
- **Sufficient insights from CI;**
 - Specific layers;
 - Limited amount of time;
- **Openness vs Secrecy;**
 - Hidden values
- **Proper classification of information;**
- **Critical Infrastructure Protection risk;**
 - Multiple areas with overlapping events
 - Possible correlation

Future work

- Research scoped to Netherlands;
 - Further cities:
 - Den Haag;
 - Rotterdam;
 - Additional dataset and sources;
- Real-time APIs;
- Automated approaches;
 - Machine learning;
 - Artificial Intelligence;
- Ongoing country profiling;
 - Business purposes;
- Open Discussion
 - Beneficial or Dangerous ?



Questions ?

