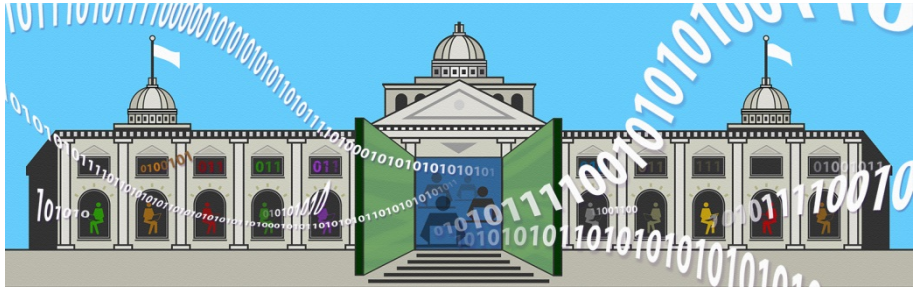
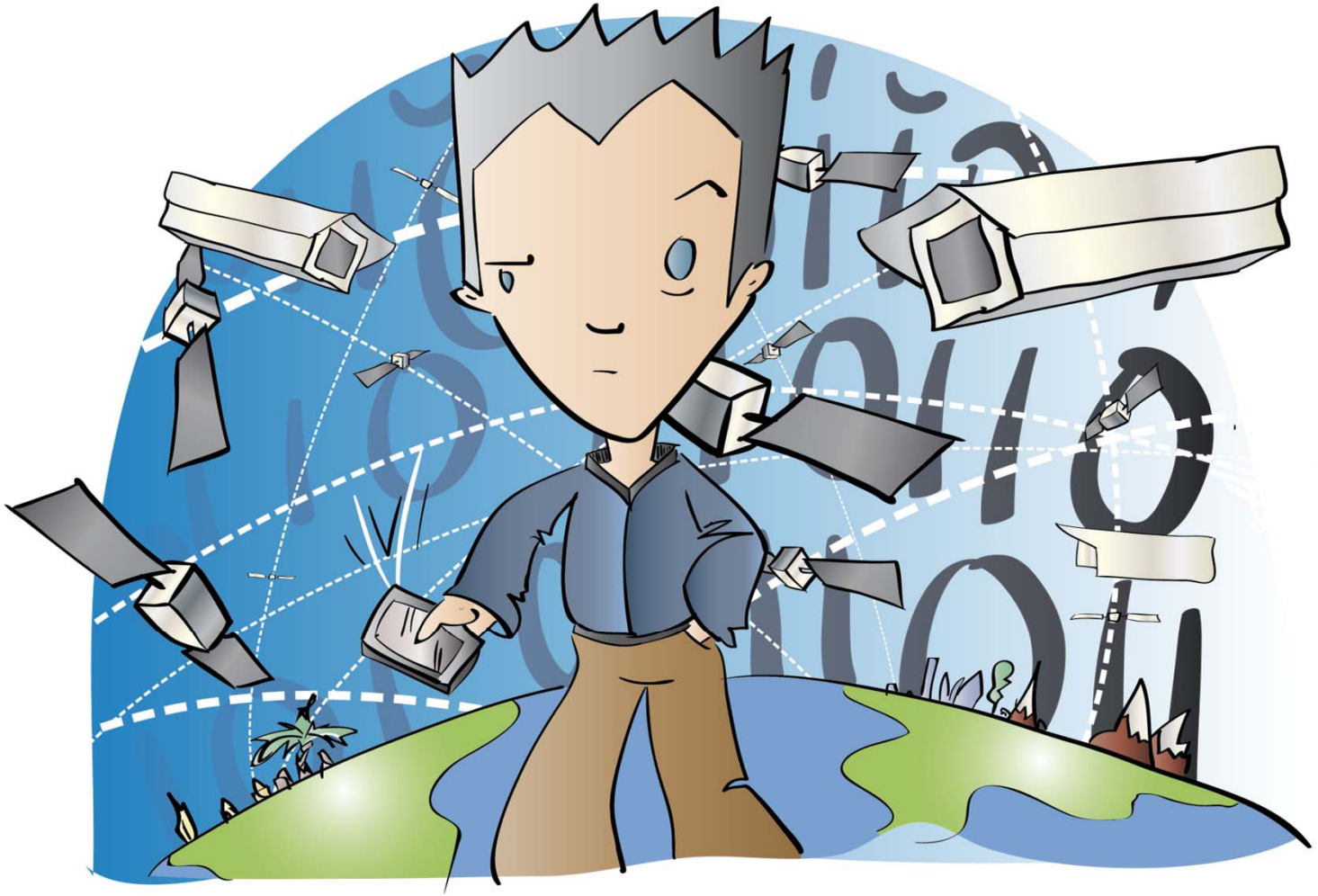
# Frontiers of big and open linked data

Seminar 11 May 2016, University of Minho, Braga

Prof.dr.ir. Marijn Janssen
Delft University of Technology

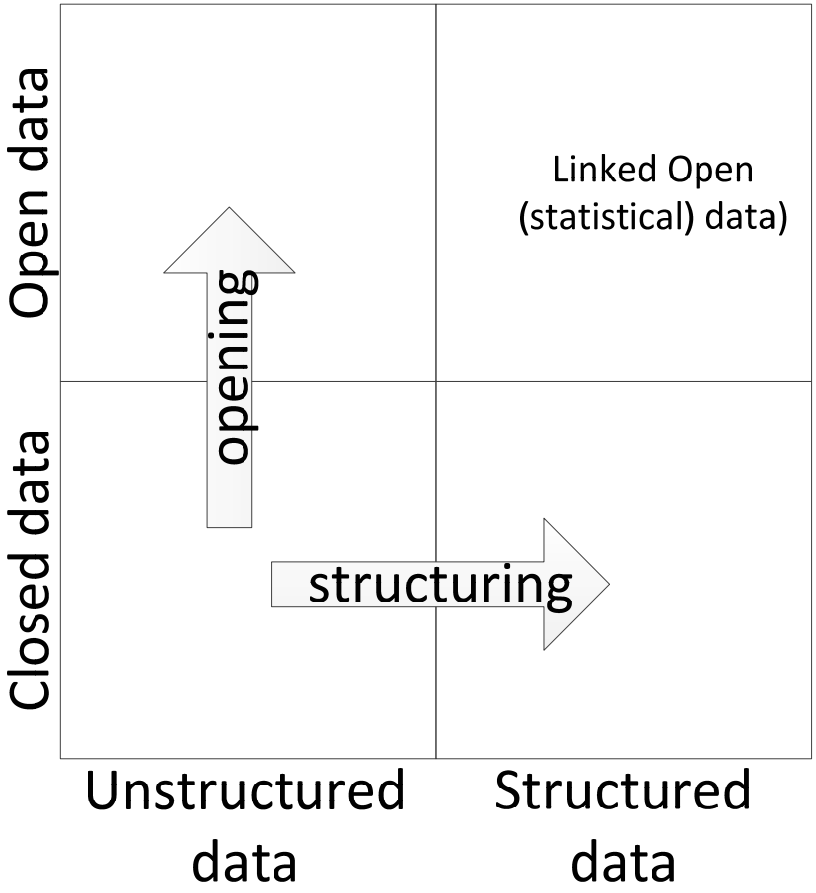


# Datification



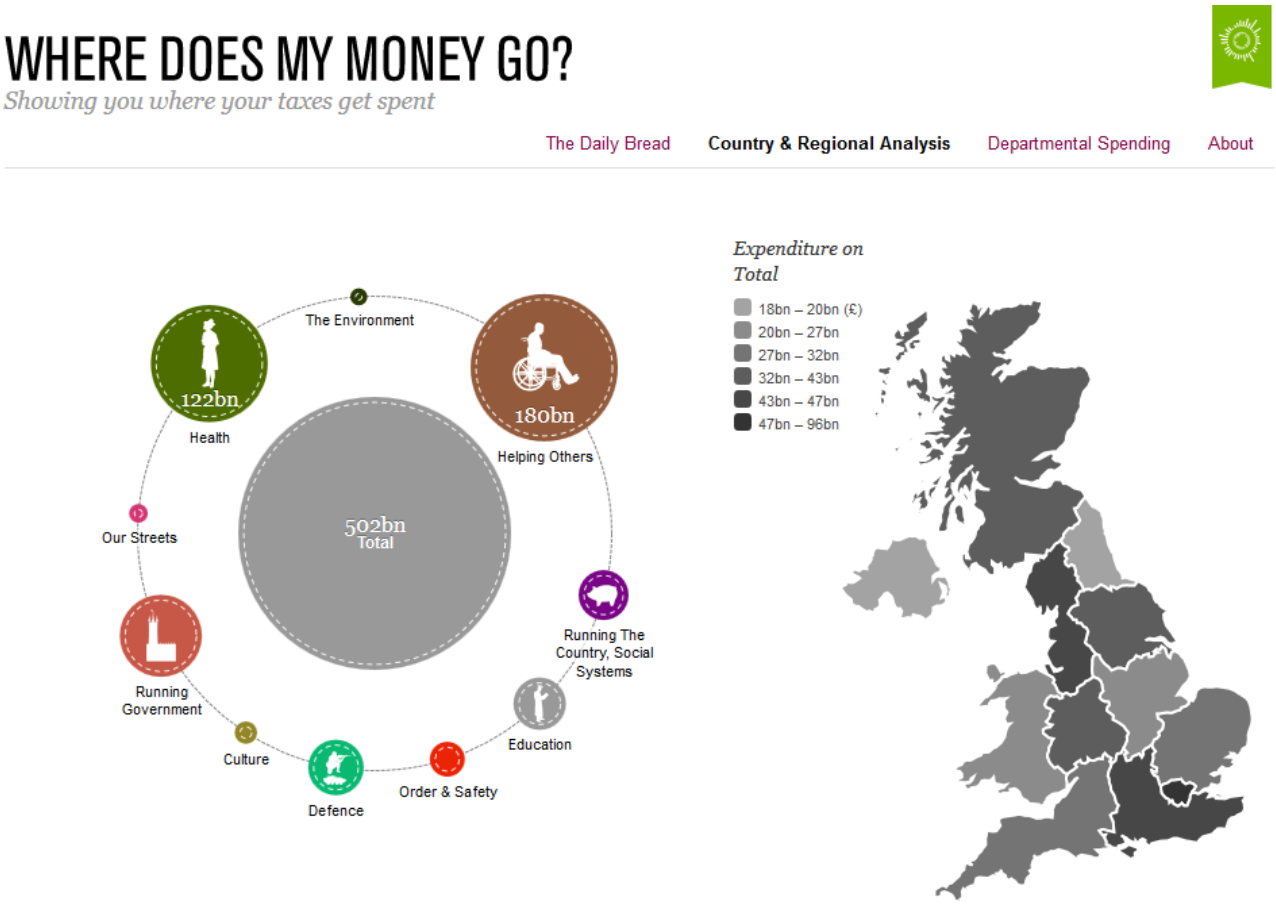
Illustrations by Annemarie van der Linde

# Developments*

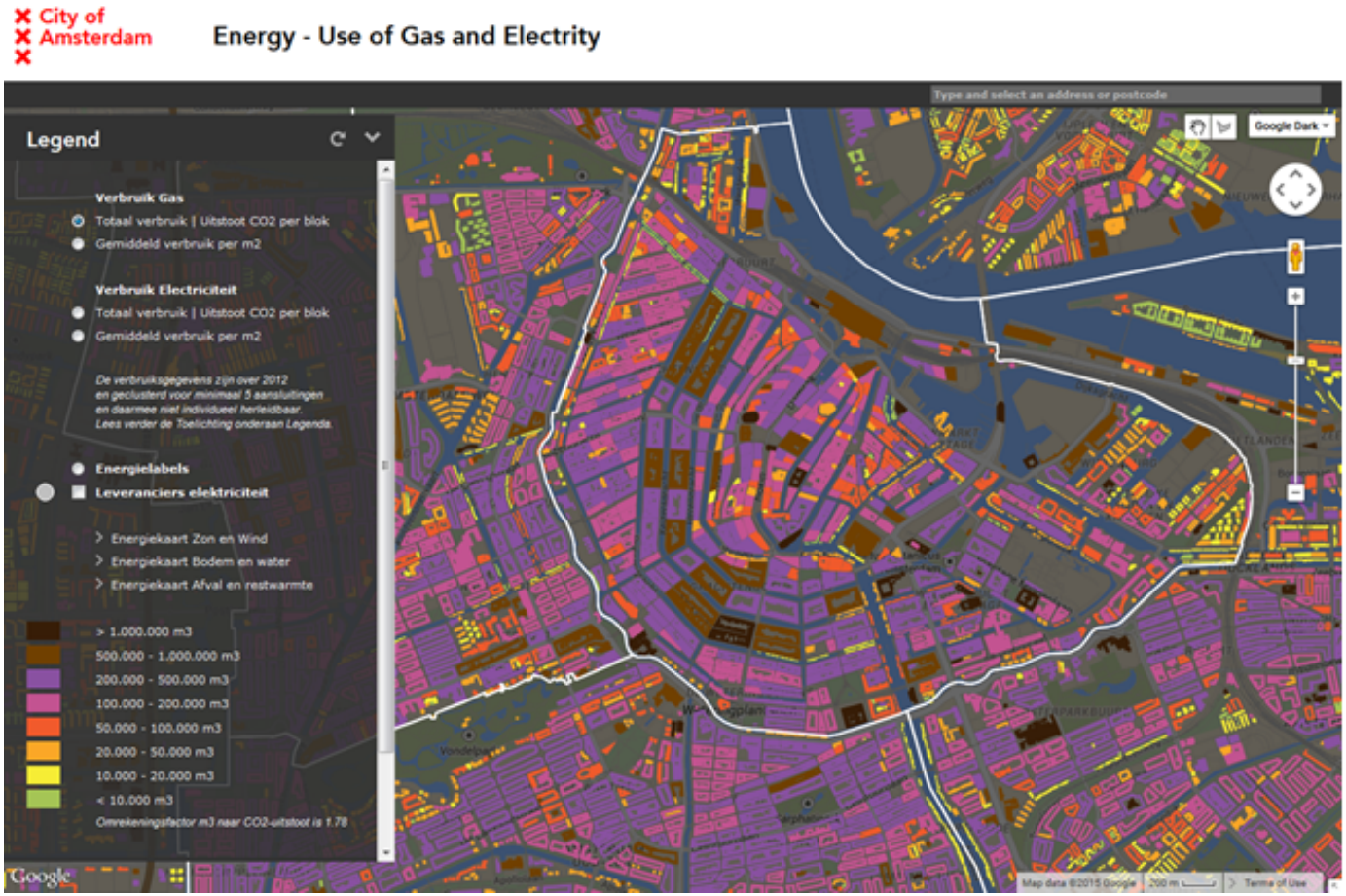



* Janssen, Marijn & Kuk, George (2016). Big and Open Linked Data (BOLD) in Research, Policy and Practice. *Journal of Organizational Computing and Electronic Commerce*, Vol. 26, no 1-2, pp. 3-13. DOI 10.1080/10919392.2015.1124005

# Open spending*: a hallmark of open government?



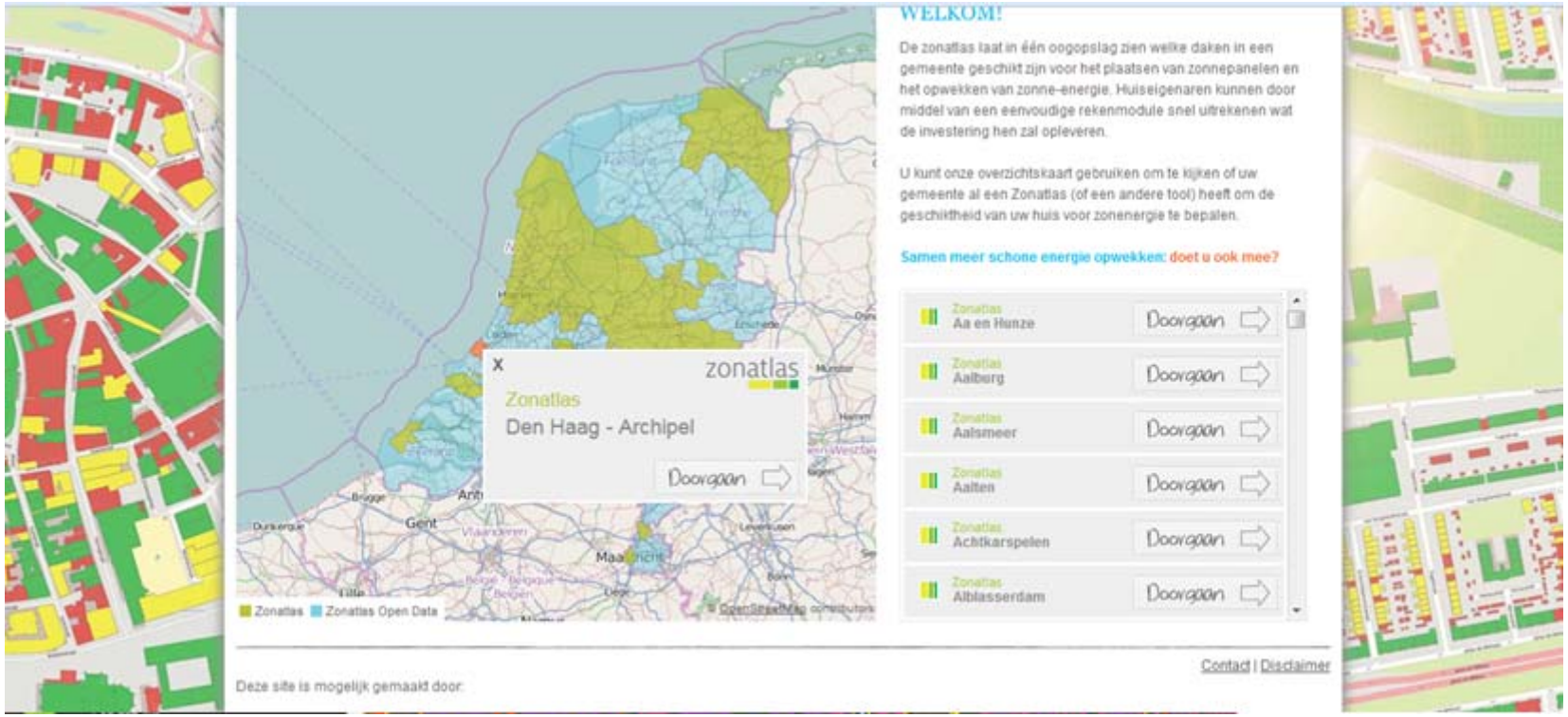
# Creating societal benefits – take societal challenges as a starting point*






* http://amsterdam.smartcityapp.nl/

# Nudging – solar cells*

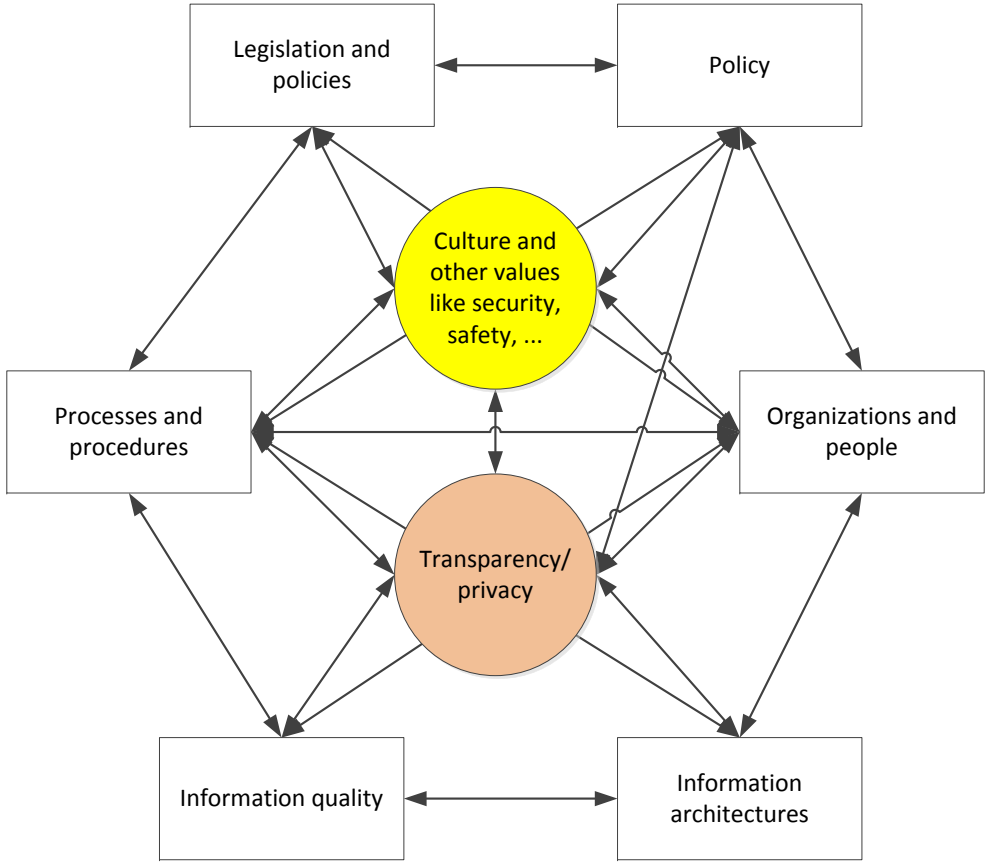


# Some challenges to create open government

- Where can we find the data?
- Who owns the data? What is the quality of the data?
- How can data be linked?
- How can the results be visualized?
- How can bias be avoided?
- Is the app user-centric? Does it address a societal problem?
- Is this creating transparency or showing a predefined view on the world?
- Who is accountable when data/results prove to be wrong/

- This is challenging
- Can we better stop? No, these efforts contribute to creating better information quality

# Realizing the benefits is not easy




*Janssen, Marijn, and Jeroen van den Hoven. "Big and Open Linked Data (BOLD) in government: A challenge to transparency and privacy?." *Government Information Quarterly* 32.4 (2015): 363-368.

# Open government and datification

- Governments are releasing their data
- The Internet of Things (IoT) is a development contributing to the collection of large amounts of data
- Greater returns from the public investment in downstream use and creation of outputs
- Most value of data is created by combining data
- Open data enables citizens and others to be involved in the policy-making process
  - By providing access to data, this data can be used by anybody to analyze the data and make suggestions for policy-improvement
  - Open data can be analyzed and the results can be used to make informed arguments for embracing, rejecting or proposing new policies
  - Transfer of activities from inside the border of the government to the outside

# Data Hugging excuses*

- It's held separately by a different organisations and we can't join it up
- It will make people angry and scared without helping them
- It is technically impossible
- We do not own the data
- The data is just too large to be published and used
- Our website cannot hold files this large
- We know the data is wrong
- We know the data is wrong, and people will tell us when it's wrong
- We know the data is wrong, and we will waste valuable resources inputting the corrections people send us
- People will draw superficial conclusions from the data without understanding the wider picture
- People will construct league tables from it
- It will generate more Freedom of Information requests
- It might be combined with other data to identify individuals/sensitive information
- It will cost too much to put it into a standard format
- Our IT suppliers will charge us a fortune to do an ad hoc extract

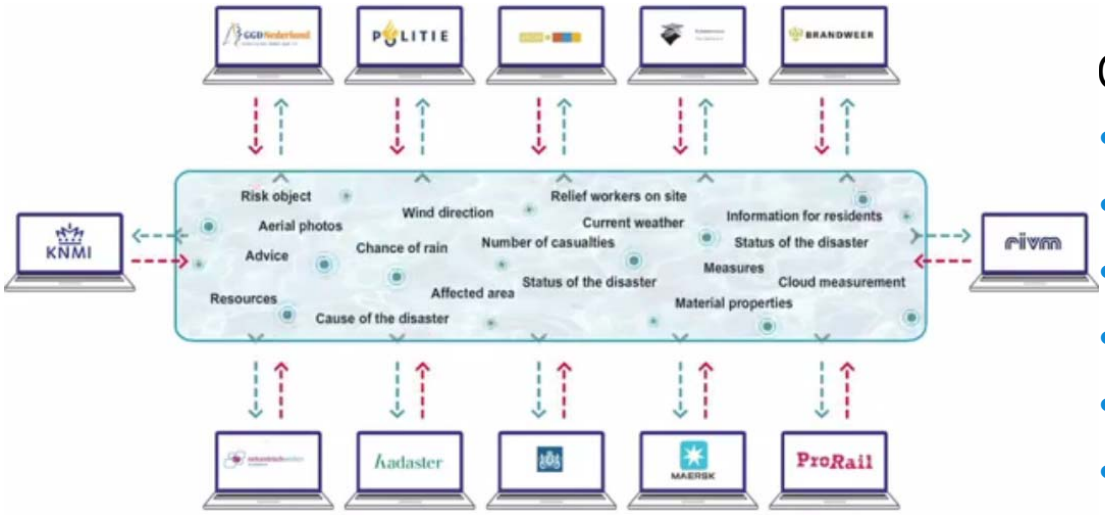
http://www.dr0i.de/lib/2011/07/04/a_sample_of_data_hugging_excuses.html

# Design principles*

| Challenges | Design principles |
|---|---|
| 1. Late involvement | Start thinking about the opening of data at the beginning of the process |
| 2. Lack of guidelines for publishing open data | Develop guidelines, especially about privacy and policy sensitivity of data |
| 3. Lack of insight in activities of other actors | Provide decision support by building in insight in the activities of other actors involved in the publishing process |
| 4. Different approaches | Make data publication an integral, well-defined and standardized part of daily procedures and routines |
| 5. Lack of focus on outcomes (e.g. data use) | Monitor how the published data are reused |

* Anneke Zuiderwijk, Marijn Janssen, Sunil Choenni & Ronald Meijer (2014). Design principles for improving the process of publishing open data, *Transforming Government: People, Process and Policy (TGPPP)*, Vol. 8 No.: 2, pp.185 – 204. DOI 10.1108/TG-07-2013-0024

# Data pool

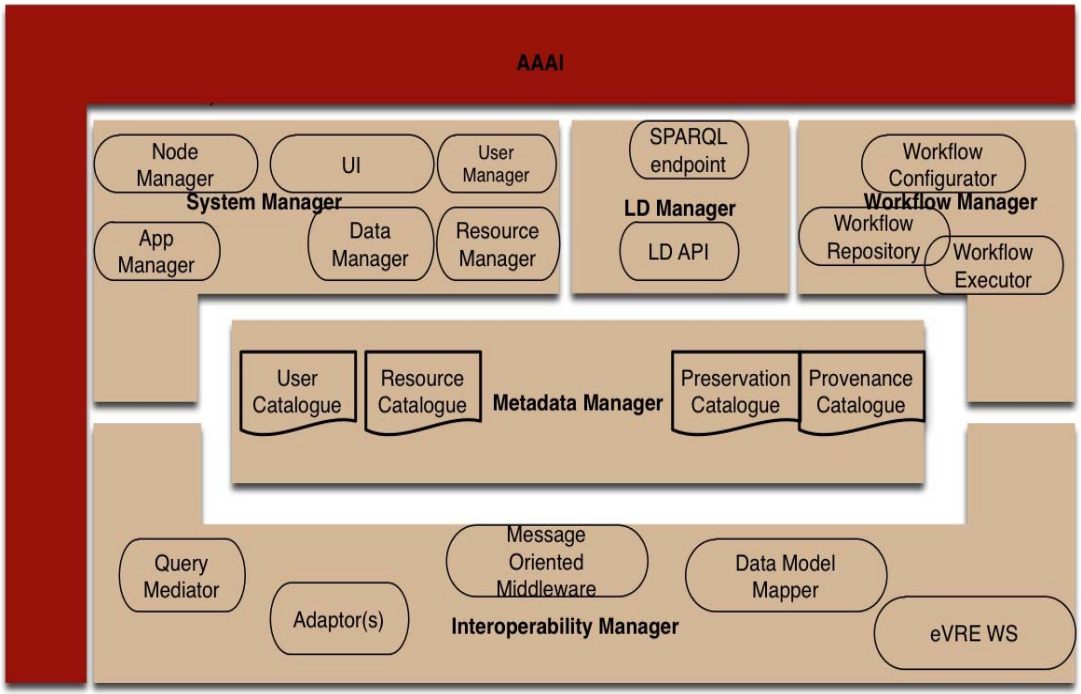



Under what conditions do you want to share data?

Conditions

- Do you trust the data?
- Do you trust the users?
- How to ensure data privacy?
- How to prevent misuse?
- What is the data quality?
- Who is accountable for (mis)use and wrong interpretation?
- ...

A Europe-wide Interoperable Virtual Research Environment to Empower Multidisciplinary Research Communities and Accelerate Innovation and Collaboration

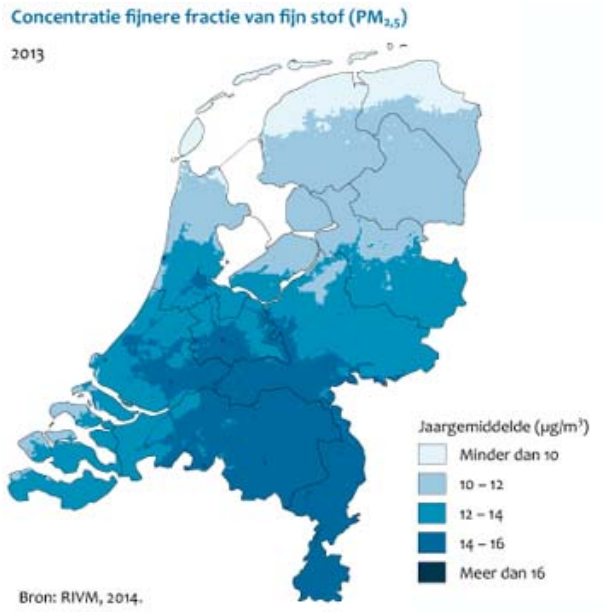



- User profiles
- Privacy-by-design
- Trust-based information sharing mechanisms

# Who leads open government?*

- People are concerned about the air quality
- Traditionally this is measured and estimated using simulations, but what is the real value in your neighbourhood?
- Internet of Things (IoT) enables low-cost measurements
- Citizens measure the air quality
- Citizens and companies design apps

- Is open government owned by governments?

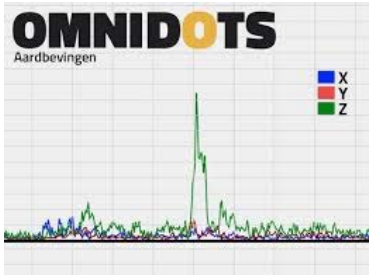


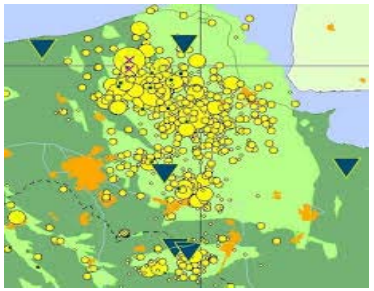



*https://www.atlasleefomgeving.nl/

# Example: Self-organization

- Earthquakes in the north of the Netherlands due to extracting natural gas
- Elected officials and policy-makers initially denied and then ignored the evidence about the impact
- Citizen sentiment turned to disappointment and unhappiness
- Citizen network to measure activity - seismometer and install it on a wall in their house
- Government  focused on compensating the costs of damage, however, the ‘real’ concern is the fear of earthquakes and unfair treatment



# Example: Self-organization*




*http://www.groninger-bodem-beweging.nl/

# Open data ecosystem puzzle

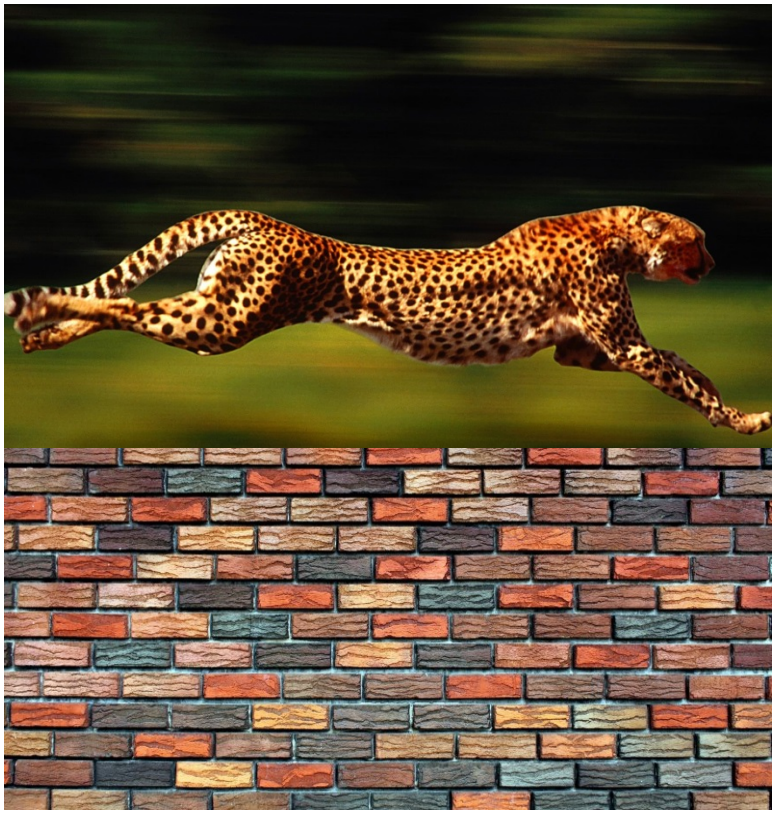
- Who owns and who has the pieces?
- Shared interest?
- Who has the capabilities?
- Competitive advantage?
- Improve the government?
- Interoperability
- ...

# Creating open government – *Government as a Platform (GaaP)**

Separate infrastructure and user-development

- Agile development (uncertainty, flexibility, user-centric)

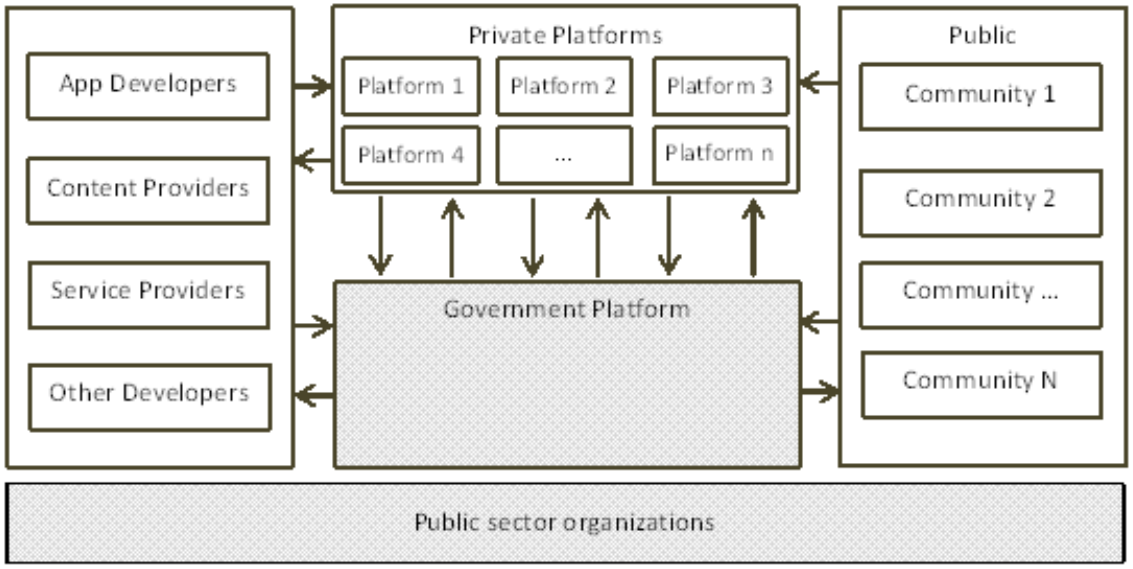
- Solid infrastructure (reuse, reliable, available, ..)



# Have the infrastructure

- Digital infrastructure is the foundation
- Usable for a wide range of (not known) opportunities
- Facilitates flexibility
- Contains readable available functions like eID, payment, security, storage, visualization, …
- Use APIs – Application Programming Interfaces – for creating flexibility
  - API – to install a new server
  - API – to log in
  - API – to access to meta-data
  - API – to access to user preferences
  - API – to visualize on a geographic map
  - …
- and designers can focus on the user-experience

# Platforms – who controls?*




- Platforms are focal points where various types of actors engage in a common environment
- People can create their own applications and can contribute with information about what is happening from multiple devices

*Source picture: Elsa Estevez & M. Janssen (2013). Lean government and platform-based governance: Doing more with Less. *Government Information Quarterly*. Vol. 30. Supplement 1, pp. S1-S8,

# Have the data

- Base registries: *trusted and authentic sources of information owned by one entity and facilitating reuse*
- Create data portfolios
- Data portfolios provide overview of sources of data
- The ability to effectively and efficiently combine, link and share data will determine the value
- Diverse set of capabilities needed
- Technical, syntactic, semantic and pragmatic interoperability

- and designers can focus on the user-experience

Table 1. Interoperability maturity stages for big and open data.

| | Stage 0: Independent | Stage 1: Ad hoc | Stage 2: Collaborative | Stage 3: Integrated | Stage 4: Unified |
|---|---|---|---|---|---|
| **Description** | An organization has no strategy and does not think about the impact of releasing or sharing its data. | Ad hoc arrangements are agreed upon for data use and interoperability within an organization. | Strategies about release and use of open data are identified, and roles and responsibilities are defined within an organization. | An organization implements shared goals and value systems with another organization based on common understandings and mutual desire for data interoperability. | An organization defines strategy, shared organizational goals, value systems, data portfolios, and knowledge bases. |
| **Levels of interoperability supported** | Technical | Technical and limited syntactic | Technical and syntactic | Technical, syntactic, and semantic | Technical, syntactic, semantic, and pragmatic |
| **Main characteristics** | Focus is on realizing benefits by developing certain applications; communications about data collection, linkage, and processing are unstructured and carried out through meetings, emails, and phone calls; no formal procedures or support are in place. | Information is used without formal governance procedures, standards, planning, or infrastructures; some overarching goals are in place, but a detailed strategy is lacking. | Governance mechanisms and roles and procedures are all in place for data acquisition, processing, and distribution; datasets can be related to other datasets to create value; strategy includes development of capabilities to ensure organizational readiness to interoperate. | Infrastructure, agreements with data providers, and assessments of information use and impact are in place; metadata is shared among organizations to enable linking and combining of data; data is shared in large volumes; manual processing is used for more complex operations, handling exceptions, and integration of heterogeneous data sources. | Governance for data portfolio use is defined; capabilities for the discovery, assessment, and integration of new data sources for a certain architecture within a short time frame are developed; data is viewed as an essential asset by the organization, and business value is created by quickly acquiring and using the data; data portfolio instruments are used to manage data quality, legal status, and permitted uses. |
| **Additional benefits created in comparison with previous stages** | | • Innovative data-based applications<br>• An understanding of demand | • Higher-quality applications<br>• Easier maintenance<br>• Developed networks and partnerships | • Higher efficiency<br>• Shorter development times<br>• Defined partnership networks | • Faster integration of new data sources and improved value creation<br>• Coherent, agile, and sustainable practice |

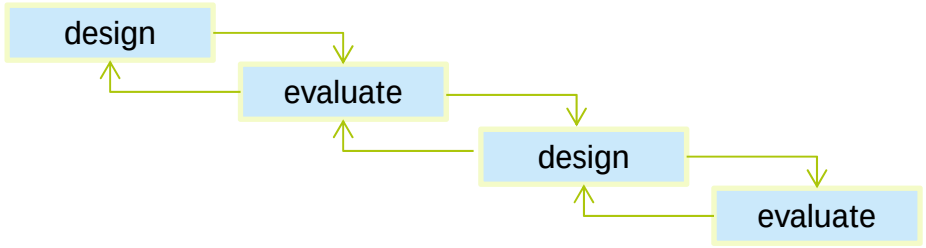


* Janssen, M., Estevez, E. and Janowski, T. (2014). Interoperability in Big, Open, and Linked Data--Organizational Maturity, Capabilities, and Data Portfolios. IEEE Computer, Vol. 47, No. 10, pp. 44-49.

# Agile development

- Requirements and solutions are created through collaboration
- Problem-solving process
- Dealing with uncertainty
- Infrastructure should enable short developments times

- Some principles
  1. Make the societal challenge and value leading
  2. Involve diversity of users
  3. Deliver working software, evaluate and improve
  4. Requirements are changing
  5. Use the infrastructure
  6. ....



Source: http://wallpaperweb.org/wallpaper/animals/sleeping-leopard_57702.htm

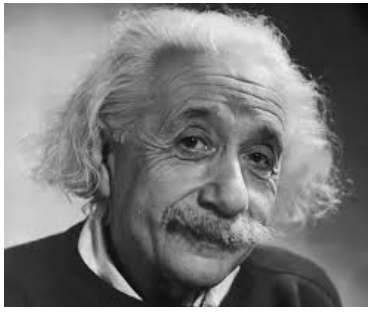
# What do we expect from users?

**Developers expect**

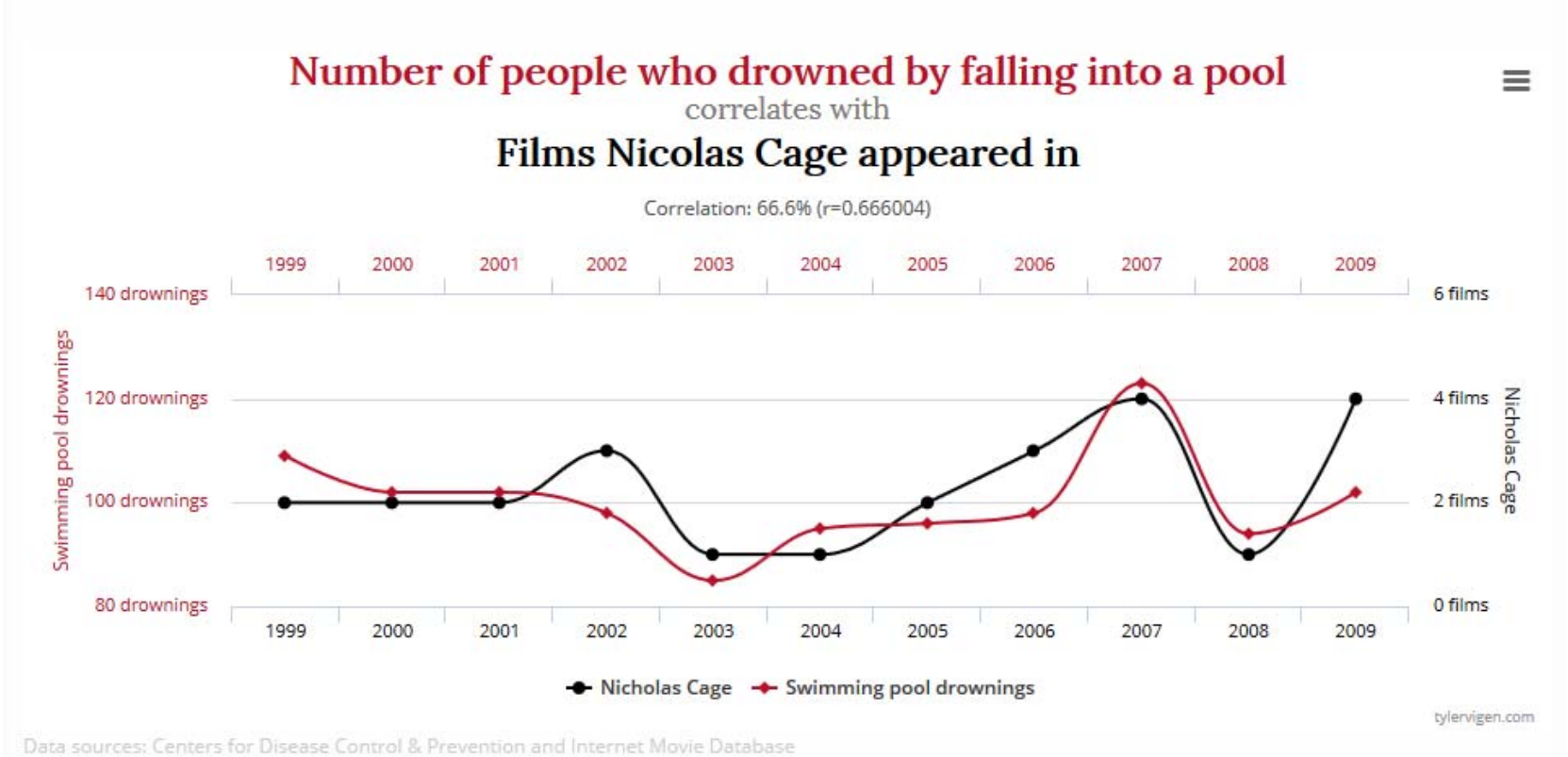
- Have the time
- Know the requirements
- Be critical
- Deep knowledge of statistics
- Contribute to social value

**In contrast users expect**

- Minimal time consumption
- Easy to understand
- No technical knowledge
- Show me my needs
- Provide value for me

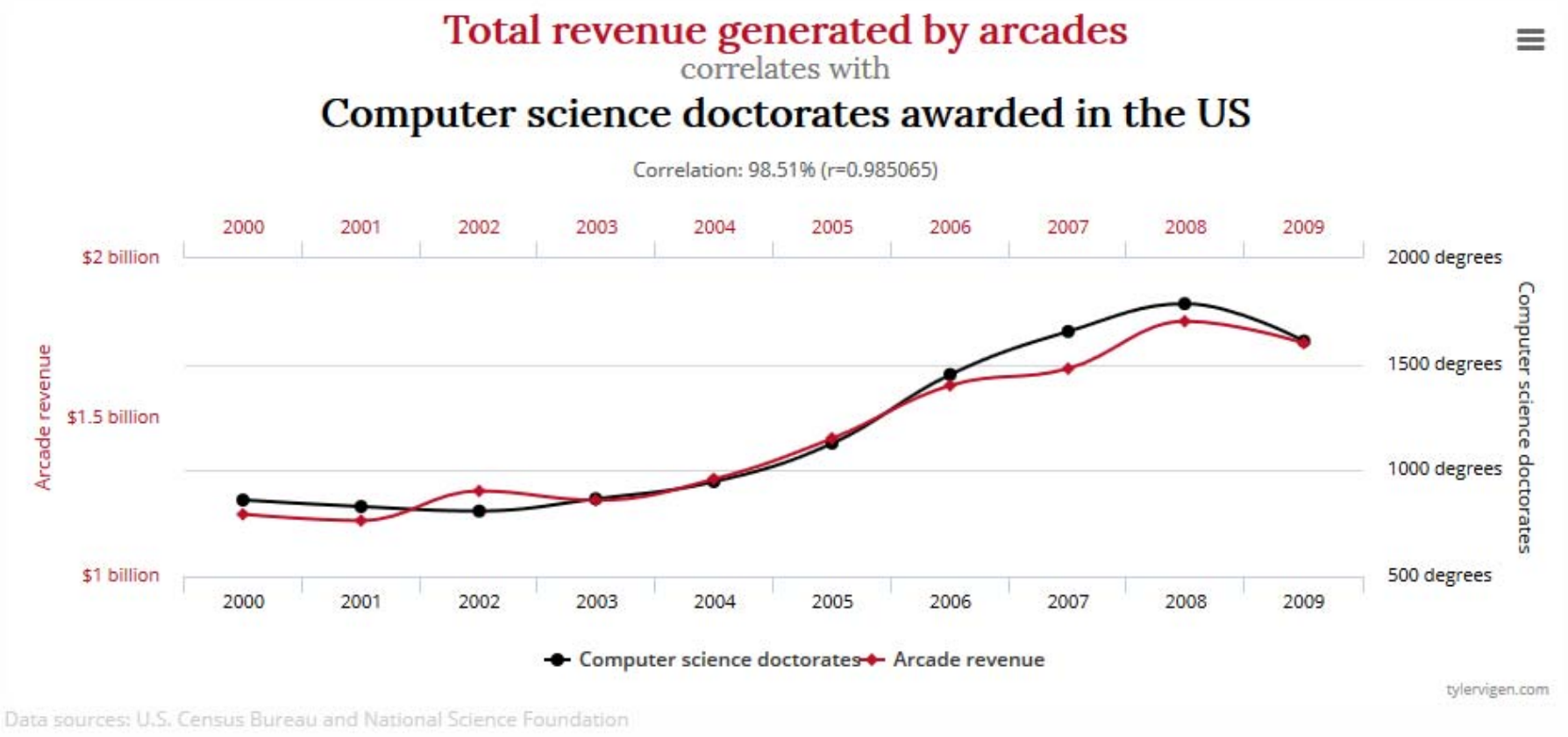


# Statistics or lies?*



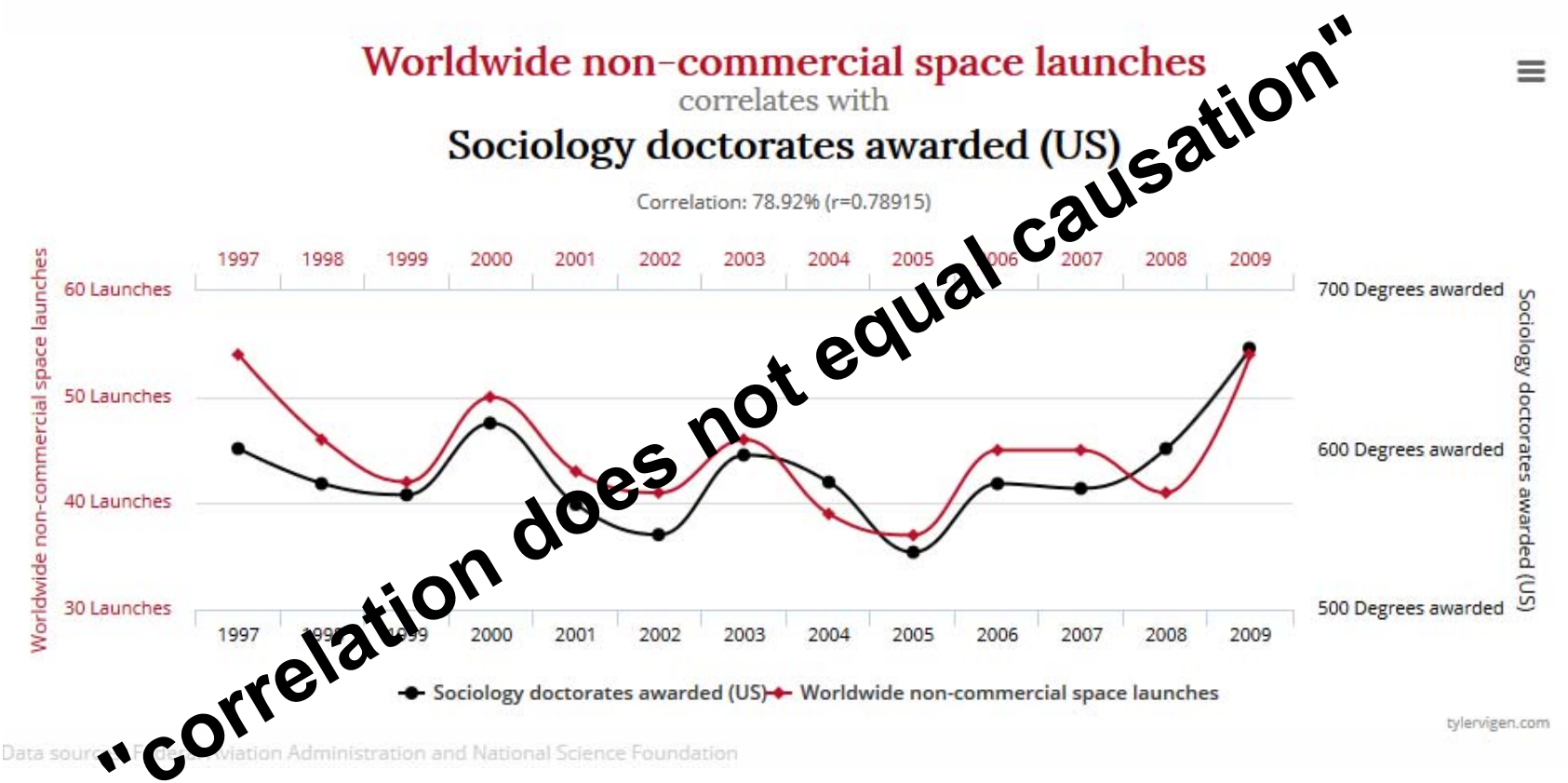

*http://tylervigen.com/spurious-correlations

# Another one



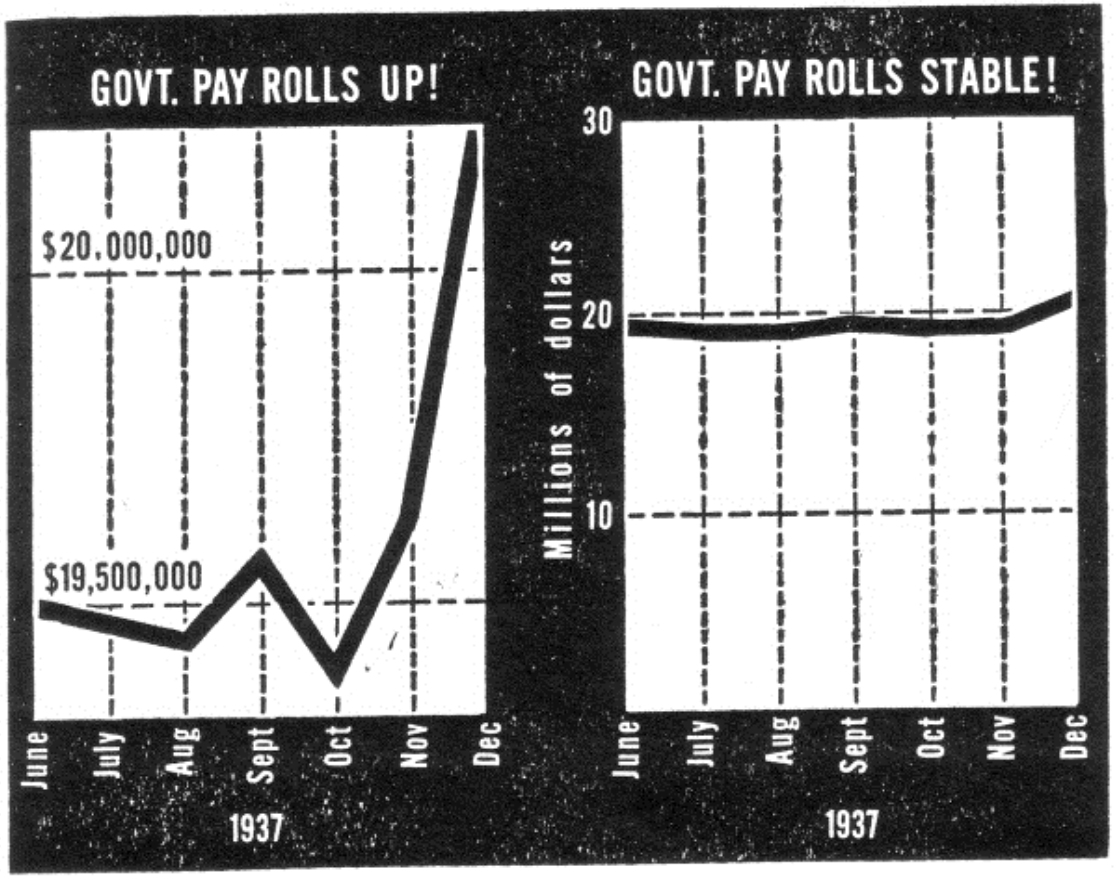

*http://tylervigen.com/spurious-correlations

# Another correlation




*http://tylervigen.com/spurious-correlations

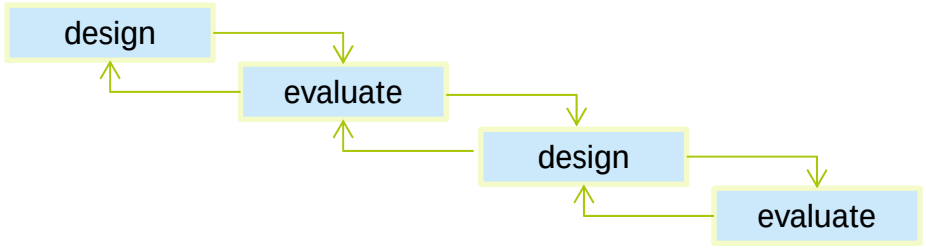
# Two ways of visualizing - what you don't want*




* Huff, D. (1993). How to lie with statistics. New York: W W Norton & Company

# Where are we heading for ..

- Open government comes from outside Focus on societal problems
- Infrastructure and data interoperability are key conditions
- Indiscriminately copying of ideas result in failure – agile development
- Not every constituent can participate and people are not uniform
- Put citizens central and not politicians (keep them out)
- Understand user behaviour by making design tangible and involving small number of users and continuously improve
- Participations does not results in transparency and trust
- Experiment and experiment ...



# Effects of open data: Bright and dark sides

# Questions?

m.f.w.h.a.janssen@tudelft.nl