Sustainability through Open Data : Examples from Switzerland

Antoine Logean[1], Oleg Lavrovsky[2], Peter Gassner[3], Ralph Straumann[4]

Abstract

Although the term “sustainability” can be traced back to as early as 1761 in forest management, it was unknown to most of the wider public two decades ago. Driven by publications of scientific evidence showing the potential impact of greenhouse gases on global climate change as well as the prediction of the end of the oil age, popular awareness of ecological challenges facing our planet has increased – today, the topic of sustainability has arrived in the mainstream (with all the downsides, such as hijacking of the term).
While awareness alone is certainly not sufficient, it is a fundamental and necessary step towards the establishment of new models for economic, societal and ecologic development. In this paper we show how the Open Data movement can address some of the complex challenges that sustainable development needs to solve. Examples of software applications (“apps”) and data analyses and visualizations developed in grassroots projects at so-called “make.opendata.ch hackathons” in Switzerland will demonstrate how the Open Data movement can be part of the solution by bridging the realms of economy, ecology and society in sustainable development.

1.Sustainability through Open Data

Since the beginning of the industrial revolution, human activity has mainly been motivated by narrow self-interest and the pursuit of short-term utility maximization [1]. Many have given less attention to the long-term impacts of (at times seemingly unlimited) economic growth. Vital natural resources such as food, biodiversity, water and energy were assumed to be virtually infinitely supplied.
In 2012, the oil production peak (“peak oil”) has almost been reached globally [2]. The current atmospheric concentration of CO_2 has increased by 30% from its pre-industrial values [3] and will continue to rise thus accelerating global change. The essentials for life to thrive (air, land, rivers and oceans) have experienced profound alterations in many places, affecting much of the Earth’s flora and fauna. The co-occurrence of these events strongly hints at the idiomatic elephant in the room: that our current model of economic development – of “doing business” – is not sustainable. Consequently, over the last 20 years numerous researchers have ventured to find out how economic growth and social needs can be balanced with the natural environment in order to ensure that the present does not adversely affect future opportunities (temporal dimension of sustainability). Meeting this challenge is incredibly difficult, as it requires global cooperation, technological innovation, changes in predominant cultural attitudes, *and* long-term commitment. In what follows we would like to show how the goals of Open Data can complement the scientific and societal efforts and how Open Data proponents can make an important contribution to the sustainability movement.

Collaboration

Sustainable development is a global issue that requires a collaborative environment to enable and enhance cooperation between different stakeholders, from corporations, local, state and federal governments, technology and service providers to the level of individual actors. Technology for communication and collabo-

[1] Antoine Logean, opendata.ch, antoine.logean@opendata.ch, @ecolix
[2] Oleg Lavrovsky, opendata.ch, oleg.lavrovsky@opendata.ch, @loleg
[3] Peter Gassner, Interactive Things, peter@interactivethings.com
[4] Ralph Straumann, ralphstraumann.ch, ralph.straumann@gmail.com, @ralphstraumann

ration plays a critical role in bridging gaps between local communities around the globe, and is a vehicle for ideas that achieve mass effect through global proliferation, most notably on social networks. Ideas build on knowledge, which in turn builds on information and data (in a slight variation of the well-known data-information-knowledge-wisdom pyramid [4]).

Sharing of ideas, knowledge, information and data has been revolutionized by the simple and robust architecture of the World Wide Web, which the Open Data movement promotes as the foundation platform for borderless data exchange. This architecture is based on the REST paradigm [5]: each dataset receives a unique identifier (a URL) [6] that can be reached from anywhere with the Hypertext Transfer Protocol (HTTP) [7]. On top of a RESTful architecture many programmable web interfaces (APIs) have been developed building what is called today the Web 2.0 [8]. These APIs enable sharing the content created in a web application with other services in a flexible way.

This has given birth to many mashups - applications created from remixes of multiple data sets by combining APIs. However, while RESTful APIs have made significant progress in making the Web a data integration platform, they are unable to provide standards in certain key areas, in particular by not addressing the issue of data classification. These problems are addressed by the Semantic Web, sometimes referred to as Web 3.0 [9]. We are particularly interested in its implementation in the form of Linked Open Data (LOD): URIs identify any kind of objects or concepts whose contents are described and retrieved using the metadata publishing and retrieval standards RDF [10] and SPARQL [11].

Like the web of hypertext, LOD is constructed with documents on the web. The difference is that, where links are relationships anchors in HTML hypertext documents, LOD defines links between data that machines can explore more effectively. The Web is no longer made only out of linked documents, but of linked datasets that are better “understood” – in the sense of connecting raw information to the underlying knowledge – both by humans and by software agents.

Complexity of the data

Sustainability is per definition a highly complex and interlinked phenomenon. This complexity stems from the scale of the interdependencies between different domains of sustainable development and the latter’s extension beyond geographic and even temporal boundaries. Being able to understand and manage dependencies efficiently requires open access to high quality, verifiable and readily available, contextual knowledge and information.

Open Data aims to open all public data, i.e. statistics and raw data about agricultural production, land cover & land use, forests, biodiversity, oceans & seas, transport & mobility, energy, atmospheric measurements, toxic chemicals, health, consumption & production, etc. Moreover, the 10 Open Data principles – completeness, primacy, timeliness, ease of physical and electronic access, machine readability, non-discrimination, use of common open standards, licensing, permanence and usage costs – as defined by the Open Government Data policy [12] ensure an exchange of information with as few barriers and impediments as possible. By using the LOD stack the complex semantic of the data can be formally specified with already existing ontologies, making their programmatic retrieval and aggregation without human intervention possible.

Data processing

Sustainability needs information-centric techniques (search, advanced analytics, integrated reporting and the like) to be able to not only keep track of complex interdependencies but to also enable objective, evidence-based decisions. These techniques are the “daily business” of the Open Data community: searching, refining, enriching, aggregating, analyzing, visualizing, disseminating and communicating data.

The central importance of technical skills such as data processing may appall some less technically inclined people. However, the technical skills per se will not make Open Data and sustainability thrive. A wider movement of people with different skillsets is necessary (as outlined above and as further detailed in the next two sections). Note, however, that one of the main assets of the Open Data movement is precisely that some people in the movement are willing to do the technical work necessary to distill data to actionable information.

Cultural change through education

Sustainability requires a major cultural change. Therefore education is critical. Through apps, analyses and visualizations, Open Data is able to show and explain complex problems to citizens in an easier-to-understand and often “hands-on” way, e.g. through *interactive* tools. The inclusion of “data/information dissemination” experts, e.g. communicators, educators, journalists and the like, is certainly beneficial to the intended public and educational outreach of the Open Data and sustainability movements. Only an educated public is capable of making sense of data and taking the right political and personal decisions; conversely, having Open Data is often a precondition to being able to educate the public on the most pressing issues.

On the other hand it becomes clear from the make.opendata.ch hackdays and from the projects subsequently described, that Open Data requires (self-)education of its own community. As some of the authors of the three showcased projects explain, they had to first educate themselves on both the subject matter and the data structures, before they could think of building something based on Open Data.

Innovation

Many sustainability problems are still awaiting solutions. Innovation and creativity are needed to imagine new solutions. We would argue that Open Data can be a major driver of innovation: through Open Data people can be newly exposed to information they have not been aware of before. Building on such freshly-acquired data, information and knowledge, they can explore potential solutions to pressing issues.

Even Open Data itself can foster innovation, mainly in the technological and methodological dimension: the afore-mentioned developments around Semantic Web / Web 3.0 technologies, Linked Open Data, metadata standards, cloud computing and “Big Data”, Service-Oriented Architectures thrive in the same ecosystem as the Open Data movement.

Furthermore, the experience from make.opendata.ch hackdays shows that Open Data attracts a rather diverse set of people, willing to collaborate on common projects. Most promising is the combination of domain experts with specialists in computer and web technology, designers and communicators. The diversity of people does justice to the highly interdisciplinary topic of sustainability mentioned above. In this way, Open Data hackdays can be seen as networking platforms and as innovation incubators or think tanks, where the best ideas are crowd-sourced from usually highly educated people, working for free with lots of enthusiasm.

In what follows, the Open (Government) Data movement in Switzerland is briefly portrayed and three Open Data projects are showcased, each of which in their own way contributing to furthering the sustainability movement.

2.The Open Government Data movement in Switzerland

Switzerland is often cited as a model democracy. It does indeed live one of the most participatory systems (what is termed direct democracy), but, most notably, does not have any substantial Open Government Data policy. Things are slowly starting to change, however. In 2010, a working group named *Opendata.ch* was founded. Over the last two years the group has been raising awareness, building alliances and considerable political leverage with regards to the cause of Open Data, resulting in a number of parliamentary inquiries on the topic. Besides, work on the ground was carried out: a significant number of local events have been held, well-attended by people as diverse as developers, designers, policy makers and business

leaders.
The following projects are centered around Open Data in Switzerland; they are the practical results of the political background work and – each in their own respect – are apt to emphasize the potential for synergy between sustainability and Open Data.

3. Exemplary Open Data projects with sustainability aspects

3.1 Swiss Army Contaminated Sites Visualization

“Swiss Army Contaminated Sites” [13-14] is an interactive tool that lets users view all officially known and registered sites in Switzerland that have been polluted through military activity, mostly through target practice (Figure 1). The tool was created during the make.opendata.ch hackdays in fall of 2011. One intention of the project was to show that opening up government data to easier access can lead to valuable insights and usefulness that go beyond that of data tables filed away in a government locker.

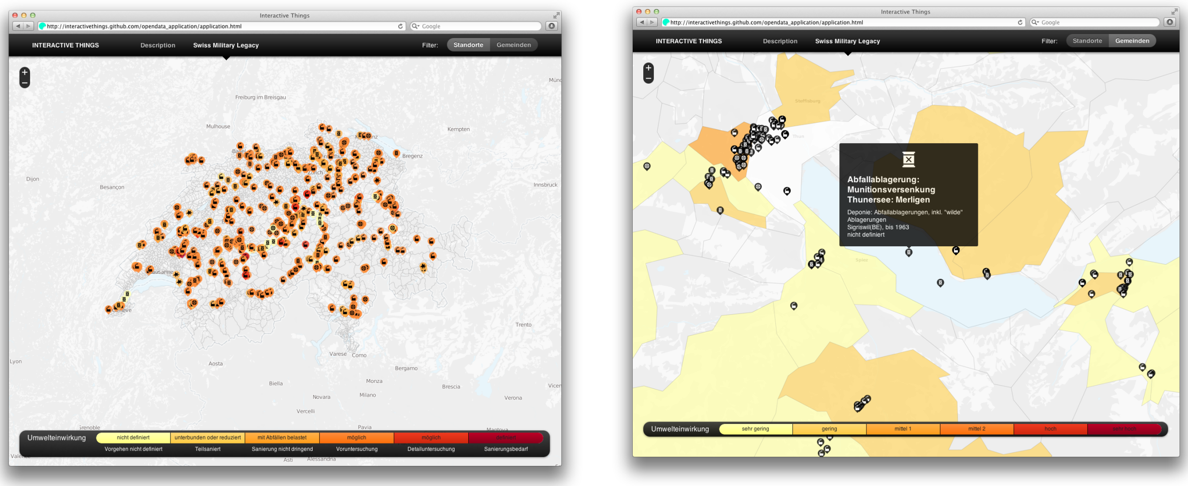

Figure 1: Screen captures of the web application *Swiss Army Contaminated Sites* [13]

During the course of the project, the authors learned that data that is *technically* accessible is not per se *easily* accessible. Even though the data had already been collected and provided to the authors by Martin Stoll of the Swiss newspaper SonntagsZeitung (for which he wrote an article based on this data [15]), a first hurdle in accessing the data were the sites’ coordinates, which were based on the Swiss geodetic system CH1903 or Swissgrid and had to be transformed to the more commonly used world geodetic system (WGS84). Another problem was language: some attributes, even though they meant the same thing, were sometimes written in French, sometimes in German. This may be acceptable for a human reader in a governmental agency, but it makes re-using the data in a software application ever so harder. All technical issues in creating this tool aside: the data itself was buried in a complex Excel table, in numbers and codes. To make the data accessible to a broader public the authors decided to create a visualization.

The authors chose to plot all known contaminated sites on a map, because this lets a user easily see patterns and regions that are more polluted than others. But, more importantly, this kind of visualization also lets users connect with the data more easily. One of the first places on a map everybody will look at is their home; it can have a strong effect when somebody sees that there is, sadly, a polluted site in their vicinity. Once this connection is made, readers can explore other aspects of the data, read what kind of site it is and how severe the pollution is. People will start asking questions and further interrogate the data, which is enabled by a carefully chosen set of icons and color codes.

The visualization also shows how much is unknown: for many sites, for example, a dump of ammunition in Lake Thun, there are no known steps to be taken. What should a reader make of this piece of information? What about polluted drinking water? What about sites we do not yet know about? The authors think that by making this data more publicly accessible, they can make readers think twice about pollution caused by military activities and the high costs associated with disposing of the polluted and polluting material at the sites.

The authors' conclusion of the project is that making governmental data more accessible to citizens – be it through newspaper articles or visualizations like the afore-mentioned – creates awareness for problems motivates people, and enables them, to ask critical questions to their representatives.

3.2 How green is my street?

Using new data provided by the IT department of the City of Lausanne at make.opendata.ch 2011, this project [16-18] compares the eco-friendliness of Lausanne's neighborhoods based on total average energy consumption for heating and electricity. The data visualization reflects the national average, as derived by researchers at the Swiss Federal Institute of Technology in Lausanne (EPFL) [19].

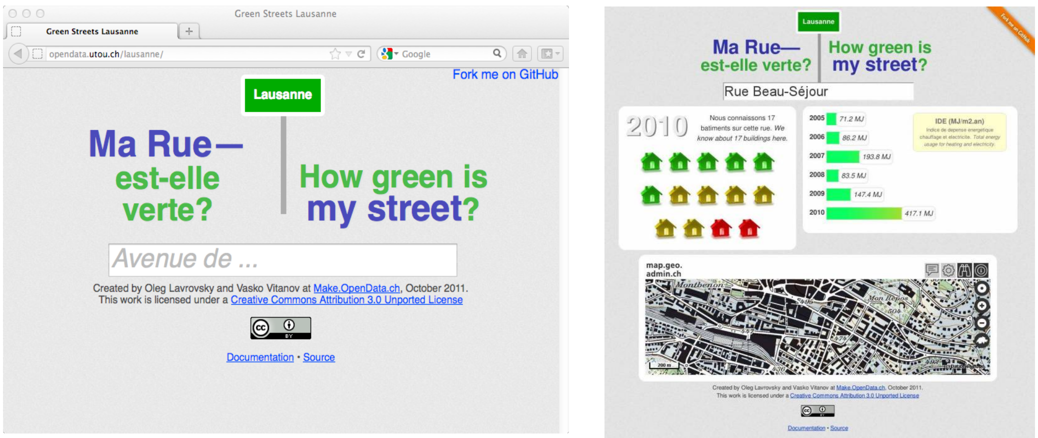

Figure 2: Screen captures of the web application *How green is my street?* [16]

The annual energy statistics for 50'000 buildings in Lausanne gathered by the town's authorities were compiled into an interactive web application where they are visualized and compared, simply by entering the name of a street (Figure 2). Once the data is loaded, the user sees a graph of annual energy consumption, as well as a map centered on the street he or she chose to investigate. Clicking on past years shows a visual representation of the building data available for that year (the number of buildings on any given street for which the data is provided varies).
Data (as well as support) was given from a representative of the Lausanne CTO attending the hackdays, as spreadsheets with over 20 columns of values for every address. The main challenges were to:
A) understand specialized data – there was a legend, but the fact that an expert and publications were available in situ was extremely helpful;

B) extract and parse the statistics outside of the original format – a conversion script in Python [20] was developed to migrate the data into an Apache CouchDB, a document-based database where MapReduce queries [21] could be used to slice and group the statistics;
C) design and develop an accessible, useful Web application which would provide limited access to the data. Another team that had imported the data into a Geographic Information System (GIS) was discouraged from publishing their results due the privacy implications of opening data on individually identifiable addresses. The authors thus decided to group buildings together into streets, never showing only one building, and used the Swiss GeoAdmin [22] map service to show the location of city streets. The application was created with HTML5, CSS3 and jQuery.

3.3 People at risk from nuclear facilities in and around Switzerland

On March 11, 2011, at 05.46 UTC a magnitude 9 earthquake about 100 km off the coast shook Japan [23]. This induced a tsunami which had devastating effects on human lives and infrastructure on the east coast of Honshu, the Japanese main island. The sequence of events is probably known to every adult of today: the Fukushima Daiichi nuclear plant experienced a catastrophic meltdown and many people had to be evacuated from the surroundings. The piece of work at hand was motivated by these devastating events in Japan and the question asked in the investigation was:

What is the number of people in Switzerland who are put at risk by nuclear facilities?

To that end, the local legal bases for emergency measures such as evacuation have been studied as well as empiric evacuation distances in the case of the catastrophe in Japan. The former were publicly available from Swiss authorities, the latter were mainly taken from international media, e.g. [24-26] (recommended evacuation distances in Japan varied widely from 20 through to 80 km). In order to assess the number of individuals at risk of evacuation (or worse) in case of a failure of either of each four nuclear facilities in and four facilities around Switzerland, a national census dataset could be obtained free of charge from the Federal Statistics Office [27]. The data specified population at 100m spatial resolution across Switzerland. Nuclear facility locations have been acquired from the Swiss legal texts and lobby organizations [28-29] and digitised in the freely available Google Earth package. A simple spatial analysis workflow was devised using Python programming language as well as a Geographic Information System (GIS). The analysis encompassed compilation of a map of nuclear facilities, radii around them and the population density as indicated by the census dataset (Figure 3, left) as well as computation of cumulative, distance-dependent distributions of individuals around the eight nuclear facilities (Figure 3, right).

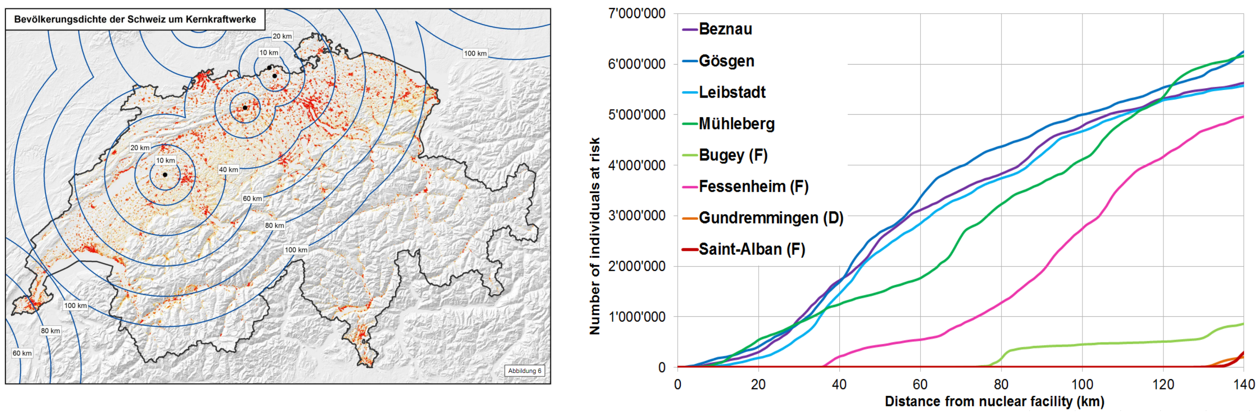

Figure 3: Population density map and locations of nuclear facilities (left), Distance-dependent cumulative distributions of individuals around nuclear facilities (right).

Switzerland being a direct democracy, initiatives and referenda are relatively easily achieved, also at a national level. Thus, it is likely that the exit out of the nuclear technology which the Swiss government have adopted in the wake of the events in Fukushima may at some not too distant point be voted upon directly or indirectly (the latter e.g. through a budgetary vote). Before this background and with the above analyses in hand, it seemed obvious not only to publish the derived information, but also to try and feed the results back into the political process. Accordingly, the results were written up into a report which is freely available under a CC-BY-NC license at [30]. Furthermore, the author sent the report to the members of the two parliamentary committees dealing with energy questions. The dissemination of the information was explicitly appreciated by some members of parliament and a key fact of the report was used in a parliamentary inquiry and a motion to the Swiss government.

From an open (government) data point of view it would have been desirable to obtain full, spatially distributed demographic detail about the population, landuse/landcover data and workforce census data. All these datasets are maintained by the Federal Statistics Office, but are available only on basis of a commercially-priced subscription. Obtaining these datasets would have allowed for a more in-depth investigation of potential damages to human lives, infrastructure, food production, evacuation requirements and some other important factors.

Nevertheless, the author of the study hopes that it will serve as informational background report in the event of a popular vote on the Swiss exit out of the nuclear technology and that informed stakeholders (in this case: the Swiss people) will further the sustainability and acceptance of political decisions in this important topic area.

4.Conclusion

In a similar way as computer-based information systems have been a driving force for productivity improvement in the last five decades [31], they can play a critical role in the establishment of sustainable development. Based on Open Data, objective and reliable information about organizational sustainability and individual environmental impact can be provided and influence individual and organizational behaviors [32]. Based on Open Source frameworks [33] applying best practice principles from modern software development, sustainable development can be fostered in complex organizations.

5.References

[1] Grober, U. Die Entdeckung der Nachhaltigkeit. Kulturgeschichte eines Begriffs. Kunstmann, Munich, Germany, 2010.

[2] Wikipedia: Peak Oil, http://en.wikipedia.org/wiki/Peak_oil.

[3] Pacala, S. and Socolow, R. Stabilization Wedges: Solving the climate problem for the next 50 years with current technologies. Science 305 (2004), 968–972.

[4] Wikipedia: DIKW, http://en.wikipedia.org/wiki/DIKW.

[5] InfoQ: A brief introduction to REST, http://www.infoq.com/articles/rest-introduction.

[6] Wikipedia: URL, http://en.wikipedia.org/wiki/Uniform_resource_locator.

[7] Wikipedia: HTTP, http://en.wikipedia.org/wiki/Http.

[8] Wikipedia: Web 2.0, http://en.wikipedia.org/wiki/Web_2.0.

[9] Scientific American: The Semantic Web, http://www.scientificamerican.com/article.cfm?id=the-semantic-web.

[10] W3C: RDF Primer, W3C Recommendation, http://www.w3.org/TR/rdf-primer.

[11] W3C: SPARQL Query Language for RDF, W3C Recommendation, http://www.w3.org/TR/rdf-sparql-query.

[12] Sunlight Foundation: Ten principles for opening up government information, http://sunlightfoundation.com/policy/documents/ten-open-data-principles.
[13] Swiss Army Contaminated Sites Visualization live site: http://lab.interactivethings.com/swiss-army-contaminated-sites.
[14] Swiss Army Contaminated Sites Visualization background information: http://make.opendata.ch/doku.php?id=project:swiss_army_contaminated_sites.
[15] Stoll M., Altlasten kosten 1 Milliarde, SonntagsZeitung, 11. April 2010.
[16] "How green is my street?" demo web site, http://opendata.utou.ch/lausanne.
[17] "How green is my street?" project wiki, http://make.opendata.ch/wiki/project:green_street.
[18] "How green is my street?" source code, http://github.com/loleg/green-street.
[19] Indice de dépense énergétique –moyenne suisse 2008, Laboratoire de systèmes Energetiques EPFL, http://bit.ly/NjUNe2.
[20] Python programming language, http://python.org.
[21] Apache CouchDB, http://couchdb.apache.org.
[22] Swiss GeoPortal, http://geo.admin.ch.
[23] U.S. Geological Survey – Did you feel it? M9.0 – Near the East Coast of Honshu, Japan. http://earthquake.usgs.gov/earthquakes/dyfi/events/us/c0001xgp/us/index.html.
[24] The New York Times (2011): Japan Nuclear Disaster Put on Par With Chernobyl. http://www.nytimes.com/2011/04/12/world/asia/12japan.html.
[25] The New York Times (2011): Japanese Officials Expand Evacuation Zone. http://www.nytimes.com/interactive/2011/04/12/world/asia/20110412-evacuation.html?ref=asia.
[26] Spiegel (2011): Sperrzone um Fukushima. Wie weit ist weit genug? http://www.spiegel.de/panorama/0,1518,754440,00.html.
[27] BFS (2011): Statistik Schweiz: Übersicht Bevölkerung. http://www.bfs.admin.ch/bfs/portal/de/index/themen/01/02/blank/key/bevoelkerungsstand.html.
[28] Global 2000 (2010): AKWs in Europa. Karte. http://maps.google.at/maps/ms?hl=de&msa=0&msid=212731046368592544065.000474b3dbfa178d4a444.
[29] Nuklearforum Schweiz (2011): nuclearplanet.ch. http://www.nuklearforum.ch/122579297964-de-index.html.
[30] Straumann, R. (2011): Gefährdung der Bevölkerung der Schweiz durch Kernkraftwerke: Eine Analyse. http://visurus.wordpress.com/2011/05/12/gefahrdung-der-bevolkerung-der-schweiz-durch-kernkraftwerke-eine-analyse.
[31] Stiroh, K. J., Information Technology and the U.S. Productivity Revival: What Do the Industry Data Say? American Economic Review 92, 5 (2002), 1559–1576.
[32] Watson R. T., Computing Ethics – An Information Strategy for Environmental Sustainability, Communications of the ACM 55, 2012.
[33] Open Sustainability, Open Source Framework for Sustainability, http://www.open-sustainability.org.