How will we interact with the #WebWeWant?
Position statement for the 5th International USEWOD Workshop: Using the Web in the Age of Data, May 31st, 2015, Portoroz, Slovenia.

Position paper

Power to the agents?! In the #WebWeWant, people will critically engage with data – and data journalism can help them want to do this

Bettina Berendt
Professor at the Department of Computer Science, KU Leuven

USEWOD's premise has been, from the outset, that the Web of Data – just like the Web as a whole – only makes sense to the extent that it is used. But who or what should do the using? As Ruben Verborgh reminds us in his USEWOD 2015 blogpost, the original vision was: "agents, which would use the Web to do things for people". And as Ruben also points out: as of now, there's not much sign of such agents.

Or is there? Max van Kleek's USEWOD 2015 blogpost recalls the strong agency that platforms in the mobile-app domain have: as orchestrators, gatekeepers, and data controllers. These platforms do things, and they do them for people: the providers of apps and platforms, their owners, their shareholders, … Van Kleek warns of the consequences that a repetition of this model would have when applied in the next big application domain of networked services: the Digital Home. And he argues that the Web (the original one, the human-readable one) would be a better model for this new domain. Thus, what's missing may not be "agents that use the Web to do things for people", but "agents that use the Web to do things for different people" – not only the economically or politically powerful, and also not only the technologically savvy.

To be fair, the vision of Semantic-Web software agents that Berners-Lee, Hendler and Lassila developed in their 2001 Scientific American article on the Semantic Web (and that Ruben refers to) is at least partially democratic in the sense that the authors talk about "consumers agents" and "producer agents". They conceptualized the latter as being "the agents of individual providers", who supply information and transaction possibilities "through their Web sites". This decentralisation is of course representative of the vision of Web and Semantic Web pioneers. It did not preview market concentration tendencies such as those described by van Kleek, and could therefore not preview the power differentials that we are observing today. One reason for this blind spot in expectations may be that technocentric visions such as the 2001 Semantic-Web software agents overlook one basic fact: **The Web and all its relatives are complex socio-technical systems** (see Ramine Tinati's USEWOD 2015 blogpost), in which software agents and human agents do things and co-determine outcomes.

So how can we create a truly democratic Web in which agents do things "for all people"? In a complex socio-technical system, there can be no easy answer to this question. But whatever the answer(s), **a new role for human users of the Web** is essential:

someone who can critically examine not only verbal arguments as in the Web of Documents, but also data, whether they are RDF LOD or structured data accessible through some special-purpose API. This role is essential for debunking myths about data being "objective", "speaking for themselves", and thereby giving us access to an unequivocal truth. Instead, rather than naïvely assume such objectivity in data, we should always critically question the origins of data, the purposes with which they were collected and processed, and the methods with which this was done.

But how can we make people *want* to engage with data in the first place? We believe that one way towards this complex and ambitious goal is via intermediaries who tell us interesting, appealing, and ultimately thought- and criticism-provoking stories about data. And we believe that today, **data journalism** is a premier type of such intermediaries, because data journalists combine the expertise of how to access, interpret, and give access to data, with the expertise of storytelling that can engage wide ranges of people.

By calling for data journalism in this comprehensive skill-set sense, are we not just shifting the problem by entrusting a new elite of highly-educated people with explaining the (data) world to us? A partial answer to this conundrum is the observation that data journalism is usually produced in teams consisting of journalists, designers, and developers (with the latter being the data experts in the computational sense). And this observation calls for **new forms of team education towards data competencies**.

In the recent 2-day **hackathon News Hack 2015**, led by Kris Vanhemelryck, we have tested our initial ideas for such an education. News Hack is part of the interdisciplinary community Nieuwslab.be, in which seven degree programmes in the KU Leuven Association collaborate via a blog, internships and workshops in order to stimulate journalistic innovation.

News Hack 2015 assembled three interdisciplinary groups of 6-7 students each from journalism, media studies, computer science, and related disciplines. Their task was to "make an innovative, digital and transmedial news story starting from a top-level question and initial datasets that you receive from the organisers. You choose the specific questions, the form, perspective and tools that you use to tell this story." After three hours of preparation, groups pitched their topics to the coaches and the plenum, and then returned to their work and further research. They were able to use state-of-the-art audiovisual tools and get the help from coaches throughout. The use of specific creativity techniques and reporting methods were strongly encouraged. All groups employed live and/or telephone interviews with citizens and other stakeholders of their stories. On the afternoon of the second day, the final products were presented to a jury consisting of journalists, scientists, and data scientists from industry. Prizes were awarded for best story, best form, best research, best data use, and highest innovativeness.

The results were impressive . They show the potential of the "pressure cooker" approach of this intense two-day collaboration. A particularly interesting aspect are the different ways in which the three teams **used – or did not use – data to engage their readers in storytelling**:

The 'Blijf van mijn lijf' editorial team ("stay away from me") reported on sexual harassment, triggered by the #wijoverdrijvenniet ("we are not exaggerating") trending hashtag that started a few days before the hackathon, used by women to describe the manifold experiences of everyday harassment they encounter. In addition to conducting background research and interviews with experts, the team extracted 2,000 tweets and processed them with natural-language techniques into a Dadaist poem. The poem is a

form of data visualisation that invites the reader to contemplate a random – and yet representative?! or totally biased?! – sample from the continuing story of this hashtag that keeps being told by Twitter users in their thousands (e.g., 18,000 tweets counted in a week for #wijoverdrijvenniet or 57,000 for the earlier German equivalent #aufschrei; more recent trending hashtags with similar topics include #dailyracism).

'Kerk te koop' ("Church for sale") investigated possible futures for the 1800 Flemish churches. Will they be remodelled as car repair shops, luxury hotels, or bookshops? The team contacted a Church Real Estate Agent, interviewed people in the street, and gave readers ways of interacting with data on church attendance in different life contexts over the past 50 years.

The editorial team of 'Reinvent Leuven' investigated the telling of migrants' stories. They created a set of mini-documentaries, each guided by one person's answer(s) to the question "What do you miss in Leuven?", and set up a Facebook Community. Interestingly, this approach (inspired by the photoblog Humans of New York) uses no data beyond some initial informational statistics, relying instead on the audiovisual potency of the individual personal story for engaging its readers/contributors.

In sum, we believe that data journalism can be a vehicle for engaging more and more diverse people into looking at, interacting with, and hopefully also questioning data. We also believe that ideas and techniques of data journalism can be taught to interdisciplinary teams in short and entertaining educational events such as the News Hack 2015 hackathon, and that such educational experiences can help shape the students' further attitudes to data in the news. (Events similar to News Hack 2015 have recently become very popular, cf. Hackastory or the ONA Student Newsroom.)

Of course, much more work is needed to substantiate these claims. Expectations regarding the extent to which media users actually do use, understand, or even question data in data journalism must be tested empirically. Alongside such evaluations, tried-and-tested approaches to good storytelling with data need to be assembled into a body of knowledge, skills and tools that further creators can work with. As one contribution towards this goal, we have started assembling and commenting on tools that we used in and around News Hack 2015.

Finally, data journalism is only one approach to making people engage with data, and we believe that the topic should ultimately find its way into more general curricula for digital literacy (see here and here for teaching examples). For computer science, the question will be how to develop next-generation software agents that meet the collaborative needs of the newly empowered human agents.