R Consortium Proposal: School of Data Material Development

The context

The field of data journalism has seen an exponential rise in activity during for the past 10 years, in conjunction with the open data movement and the multiplication of user-friendly tools aimed at journalists and other similar data users. With its annual international conferences, the growth of university classes and online courses, and a sustained funding from traditional journalism supporters and new open data funders, data journalism has become a leading topic in regard to innovation in journalism.

Defined as practice allowing journalists to combine their ability to tell a compelling story with the sheer scale and range of digital information now available, data journalism is also seen as a new approach to improve the role of media as a government watchdog. Concrete examples of the impact of data journalism are plenty (Panama Papers, Migrant Files, World Bank eviction scandal, ...) and they all reinforce the urgency to increase the adoption of the data journalism methodologies, techniques and tools.

R is one of them.

The Problem

R has great tools for accessing and analyzing open data sources, which can be hugely beneficial to journalists. Some leading newsrooms, such as ProPublica and the New York Times do use R in their data journalism workflow, to great effect. However beyond these obvious names, there is a low level of awareness of R in the journalism community. If we also take into account the low availability of specialized R learning content – especially in languages other than English – it appears that a substantial effort is needed to reach more journalists across the world.

This problem is compounded by several other factors, that we explored in a small worldwide survey that we conducted in preparation for this proposal. It appeared that, among the journalists who had never taken an R course, the main reason was the lack of learning content addressing their specific needs. Another substantial portion had never heard of R (see our summary on Github). This strengthens our main hypothesis: there is a need for learning content specially designed for journalists.

The Plan

With the help of the survey results (see our summary), we refined our initial idea of creating journalism-specific learning content into a series of learning modules organised into tracks and based on real projects. While the early introductory modules will be fully written by our team in order to fully tailor them to journalists, we plan to use the tracks as a way to curate existing R learning content available online, making them contextually useful for journalists. When necessary we will also translate this content to make it useful for non-English speakers and adapt the data examples for national contexts. All our content will be published on GitHub and the School of Data website, under Creative Commons and MIT licences.

The four languages/countries (French/France, German/Switzerland, English/Ghana and Spanish/Costa Rica) are chosen due to countries of living of the proposal writers. In the future we plan to extend the materials to other languages/countries using the School of Data network.

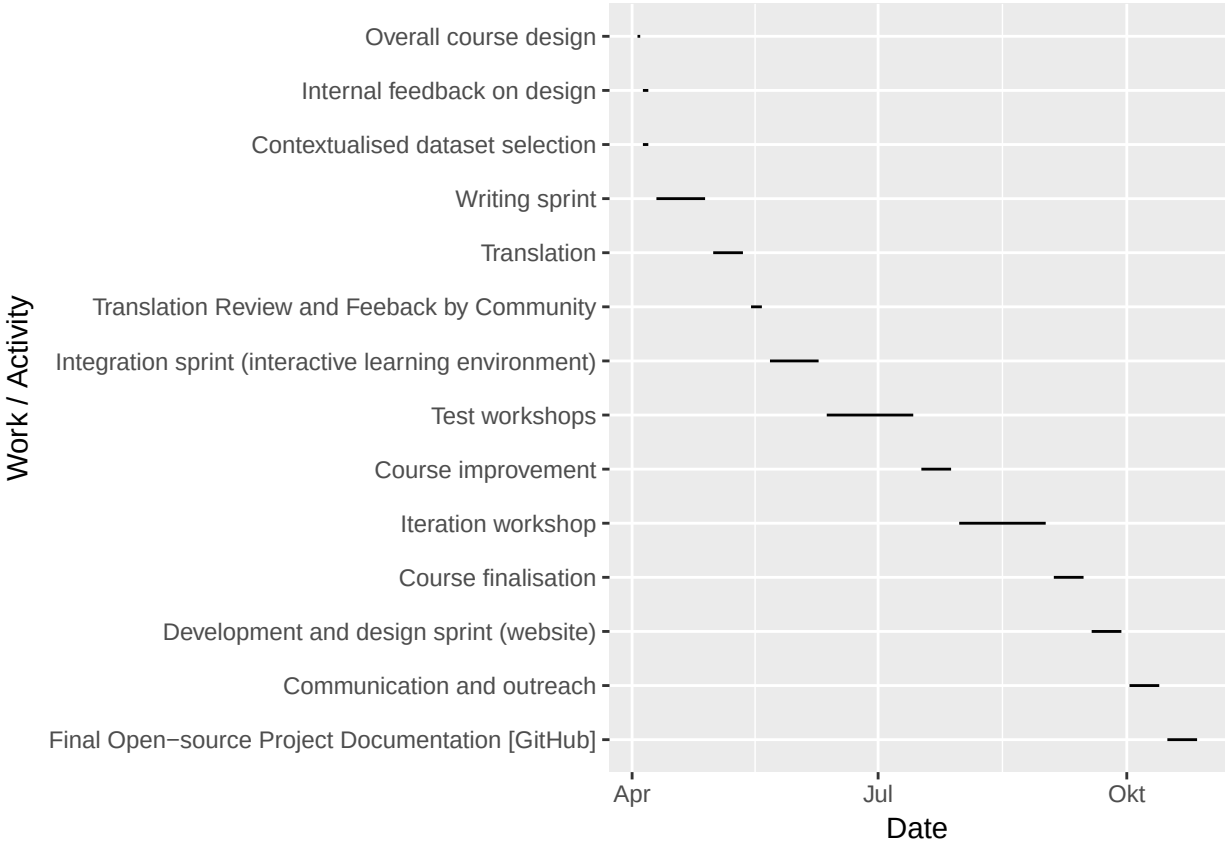
Goals

- Produce introductory learning content about R tailored for journalists in four languages: English, French, Spanish and German
- Design a content format that encourages journalists to learn without feeling overwhelmed
- Select and organize existing learning resources about R to direct journalists to them
- Optional: Organize workshops in four countries: Ghana, France, Costa Rica and Switzerland

Deliverables

- Several modules about learning R for journalists, from "why would I want to use R?" to "using R in a newsroom workflow", organised in learning tracks
- A simple GitHub-based website displaying the content in an attractive format
- Optional: Training plan for in-person workshops
- Optional: Slides for in-person workshops

Timeline



Likely failure modes and how to recover

As we will prepare the material in a modular way, it will not be a big problem if we do not manage preparing as much material as we now anticipate. In this case we will have less modules, but quality of those will be ensured. Our content being open and reusable, anyone from the School of Data community will be able to extend the content beyond our initial work.

How Can The ISC Help?

We would like to request financial support of 11200 USD from the ISC to develop material for journalists who would like to learn R. The School of Data can support the material development in guidance for the material writing as well as in communication and maintenance. The development costs can be split up as follows:

Work	Funding requested	School of Data co-funding
Writing sprint	5000	1000
Translation	1500	0
Integration sprint (interactive learning environment)	1500	0
Test workshop	1000	0
Iteration workshop	1000	0
Development and design sprint (website)	1200	0
Communication and outreach	0	1000
Website maintenance	0	1000

If additional support is possible, we would like to organize workshops in Ghana, Costa Rica, France and Switzerland. The financial support needed are 7000 USD and can be split up as follows:

Work	Funding requested	School of Data co-funding
Workshop room	800	0
Food	200	0
Communication and outreach	2000	1000
Workshop content planning	1000	0
Workshop material (slides, etc)	1000	0
Time on workshop	1000	0
Workshop documentation based on the learning content	1000	0

Dissemination

The School of Data has a broad community with many members working in journalism.

We will promote the content within and beyond our community, by sending out newsletters, tweets and information on the main School of Data website as well as the websites of the chapters in the different countries. We will publish the materials under CC-BY license and make it easy for others to build upon the content and extend it.

The team

The people

- Heidi Seibold is a PhD Student in Biostatistics in Switzerland and co-organizer of the Zurich R User Group.
- Camila Salazar, is a Costa Rican Data Journalist working at the lead newspaper La Nación.
- David Opoku, is the Africa Lead at Open Knowledge International, focused on capacity building. David is from Ghana and is a proficient R user.
- Cedric Lombion is a member of School of Data's International Coordination Team, responsible for Program Design and Implementation.
- Samuel Goëta is a French sociology researcher, data trainer and co-founder of Open Knowledge France and specialized in open data

- Joel Gombin is a French political scientist known for his detailed analysis of the political party National Front using R.

The organisation

School of Data is a network of data literacy practitioners, both organizations and individuals, implementing training and other data literacy activities in their respective countries and regions. Members of School of Data work to empower civil society organizations (CSOs), journalists, civil servants and citizens with the skills they need to use data effectively in their efforts to create better, more equitable and more sustainable societies. Over the past four years, School of Data has succeeded in developing and sustaining a thriving and active network of data literacy practitioners in partnership with our implementing partners across Europe, Latin America, Asia and Africa.

Our network includes 13 organisations across the world and a total of 101 active individuals, which all contribute to School of Data key programs: the Fellowship, the Curriculum and Member Support. Together, we have produced or are in the process of producing dozens of articles, lessons and hands-on tutorials on how to work with data. They are published on our website and are widely reused both by our network and beyond, benefiting thousands of people around the world. Additionally, we have trained over 6500 people through our tailored training events and mentored dozens of organisations to become tech savvy and data driven.

This proposal and further details are available at https://github.com/school-of-data/r-consortium-proposal