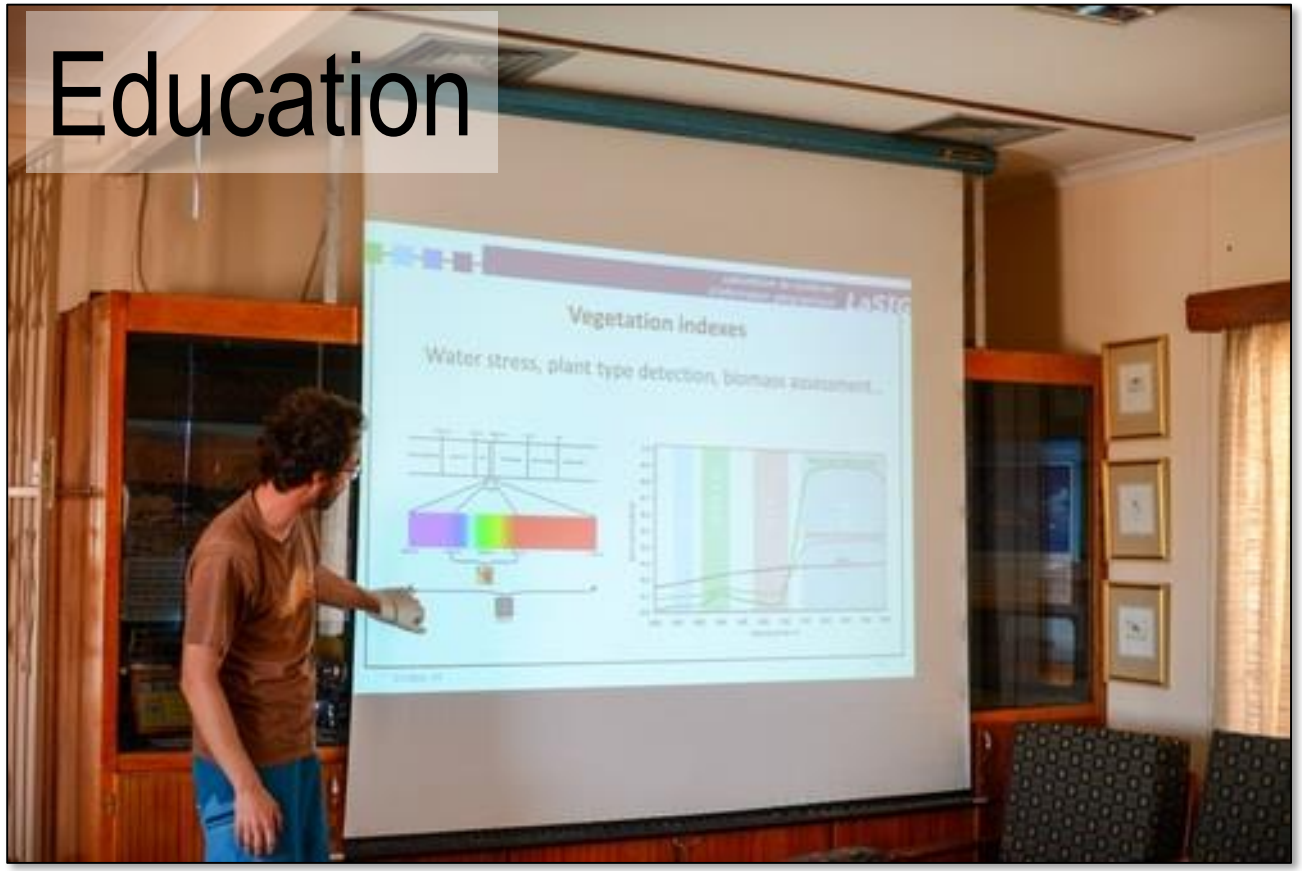
Open science practices in geospatial education and research

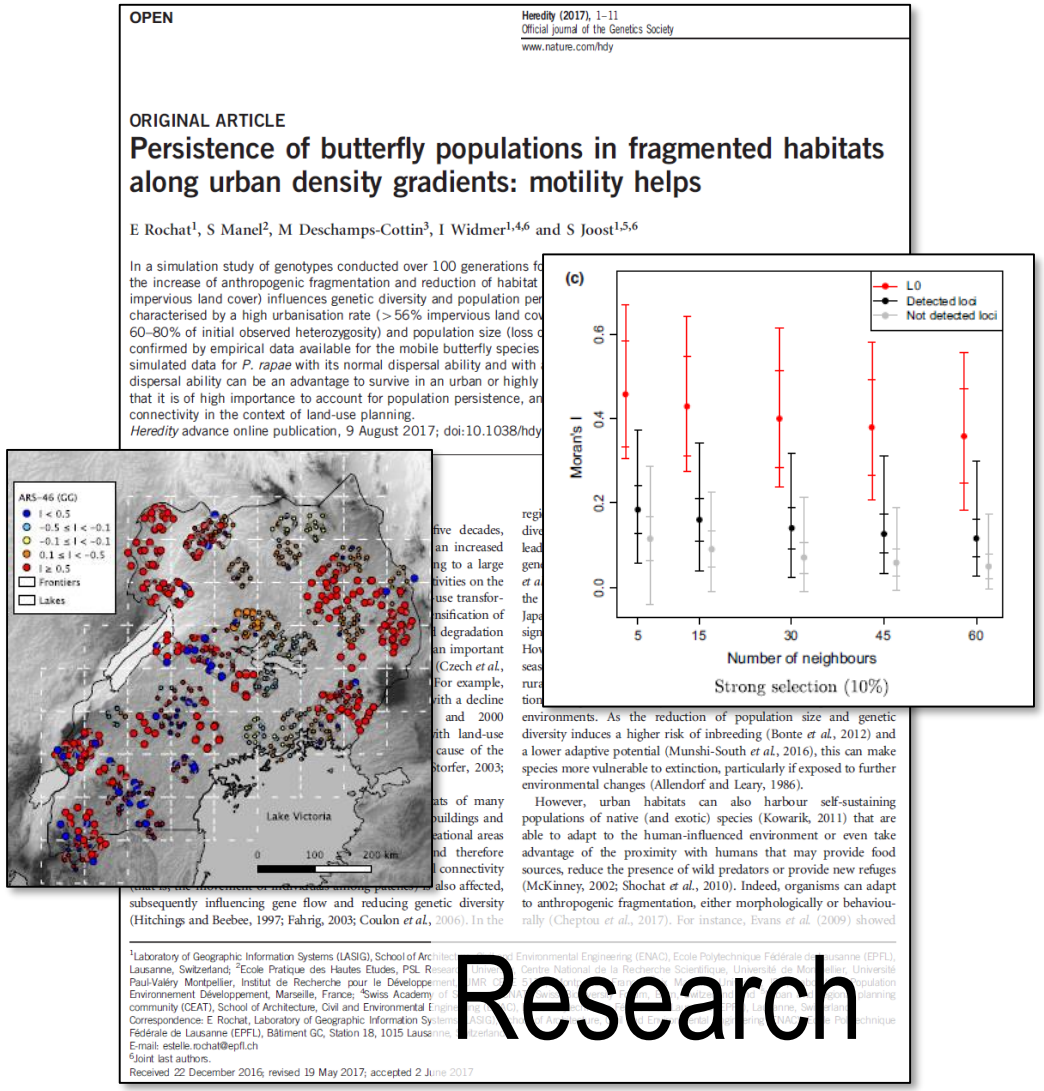
stephane.joost@epfl.ch

Laboratory of Geographic Information Systems (LASIG, EPFL)
Geographic Information Research and Analysis for Public Health (GIRAPH, EPFL&HUG)

GIS Lab – 2 missions



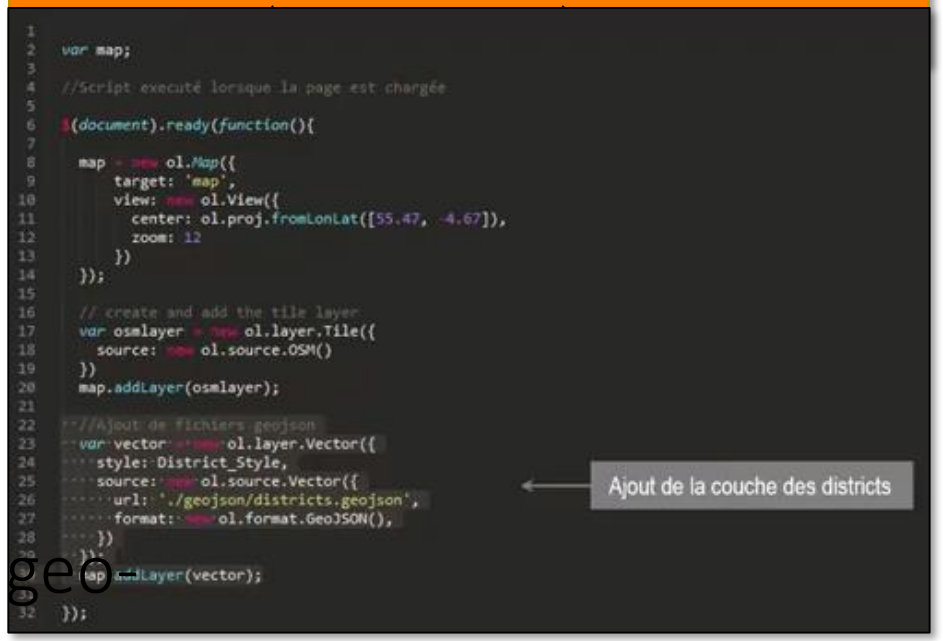
Timothée Produit (LASIG) at Gobabeb Research & Training Center, Namibia



Education

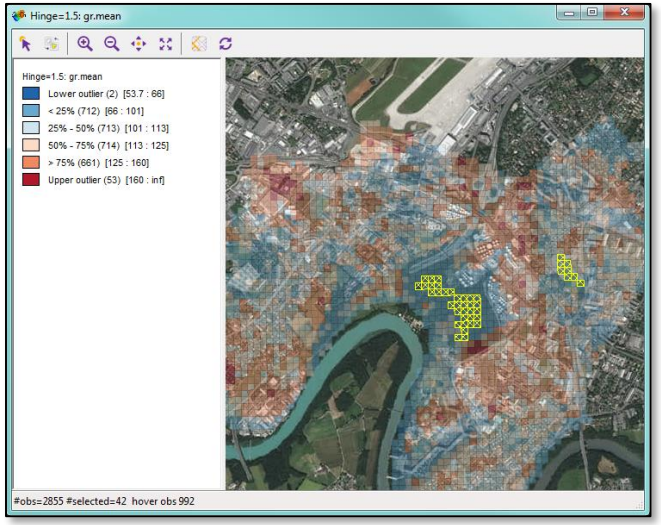
- Teach students how to use GIS software
- Avoiding Buttonology
- Empower the students by revealing the logic of algorithms
- **Open source code as text**: available for reading, manipulating, understanding
- Consider geospatial methods as tools in own research & as subjects for research
- Global situation “open source vs commercial” in geo-education is mixed
- Since 2010, GIS teaching at EPFL is exclusively based on open solutions

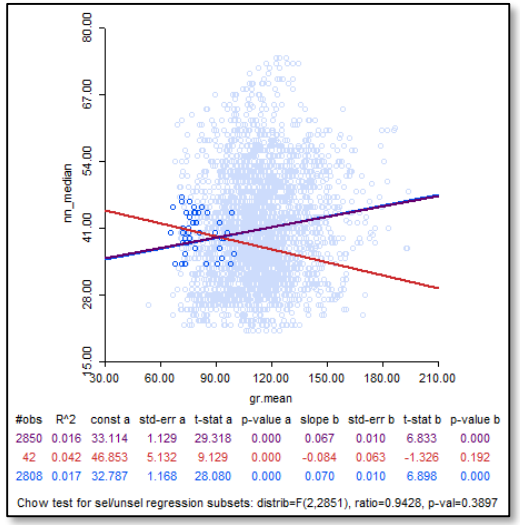
«How to use tools constrained by software licenses and negotiated by universities over

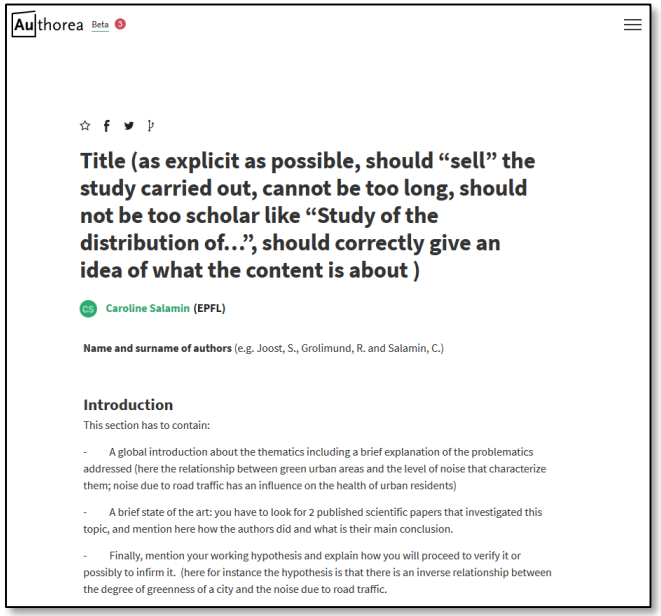


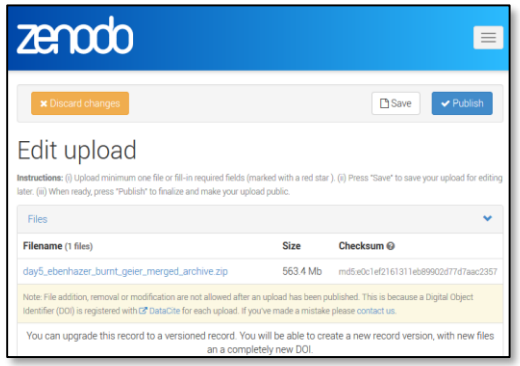
Education – collaborative writing & open access data, data repositories

- ENV-444 EDA & GVIZ
- Create an original geodataset
- Working hypothesis
- Collaborative writing with Authorea
- Manage research sources with Zotero
- Fake paper submission to Moodle
- Upload open access dataset to Zenodo Sandbox



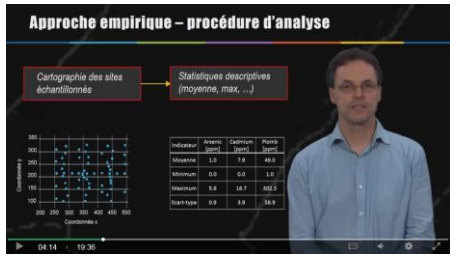




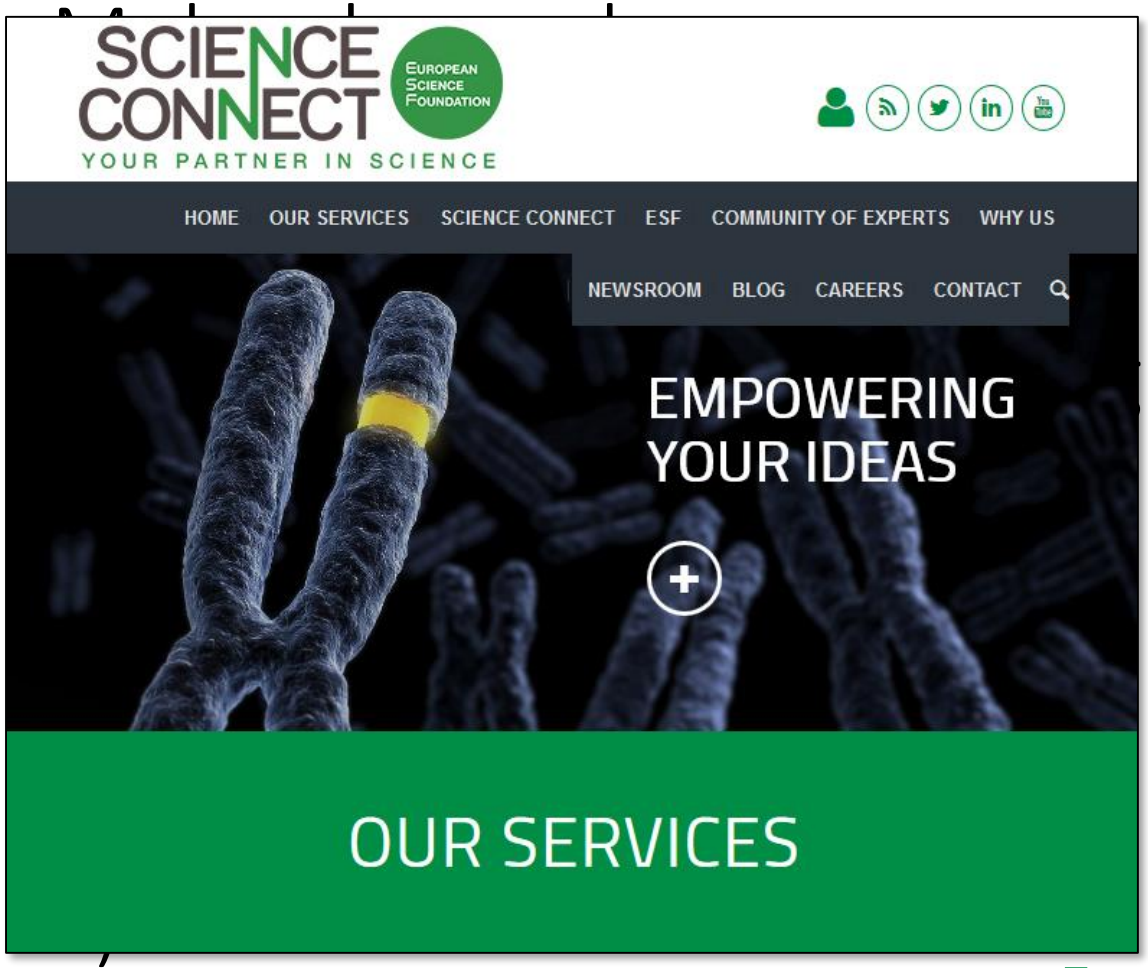



DELL

Education – 2 MOOCs, open source software only (french & english soon)



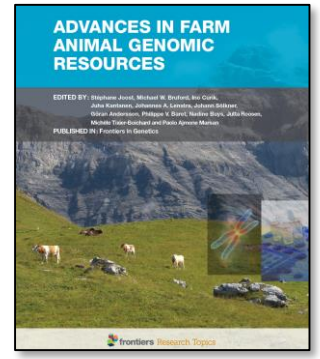
Research – open publications



Special issue – Frontiers' research topic

Articles Prices by Journal before discounts (in US$)

Journal	A Type Articles	B Type Articles	C Type Articles	D Type Articles
Frontiers in Aging Neuroscience	2,490	1,150	450	Free
Frontiers in Applied Mathematics and Statistics	1,150	700	450	Free
Frontiers in Astronomy and Space Sciences	1,150	700	450	Free
Frontiers in Behavioral Neuroscience	2,490	1,150	450	Free
Frontiers in Bioengineering and Biotechnology	1,900	875	450	Free
Frontiers in Built Environment	1,150	700	450	Free
Frontiers in Cardiovascular Medicine	1,900	875	450	Free
Frontiers in Cell and Developmental Biology	1,900	875	450	Free
Frontiers in Cellular and Infection Microbiology	2,490	1,150	450	Free
Frontiers in Cellular Neuroscience	2,490	1,150	450	Free
Frontiers in Chemistry	1,900	875	450	Free
Frontiers in Communication	950	700	450	Free
Frontiers in Computational Neuroscience	2,490	1,150	450	Free
Frontiers in Digital Humanities	950	700	450	Free
Frontiers in Earth Science	1,900	875	450	Free
Frontiers in Ecology and Evolution	1,900	875	450	Free



Project's budget to cover publishing fees

Research – software development

Open access text and code

MOLECULAR ECOLOGY RESOURCES

Explore this journal >

Open Access | Creative Commons

Resource Article

High performance computation of landscape genomic models including local indicators of spatial association

S. Stucki, P. Orozco-terWengel, B. R. Forester, S. Duruz, L. Colli, C. Masembe, R. Negrini, E. Landguth, M. R. Jones, The NEXTGEN Consortium, M. W. Bruford, P. Taberlet, S. Joost

First published: 28 November 2016 Full publication history

DOI: 10.1111/1755-0998.12629 View/save citation

Early View

Browse Early View Articles
Online Version of Record published before inclusion in an issue

nextgen

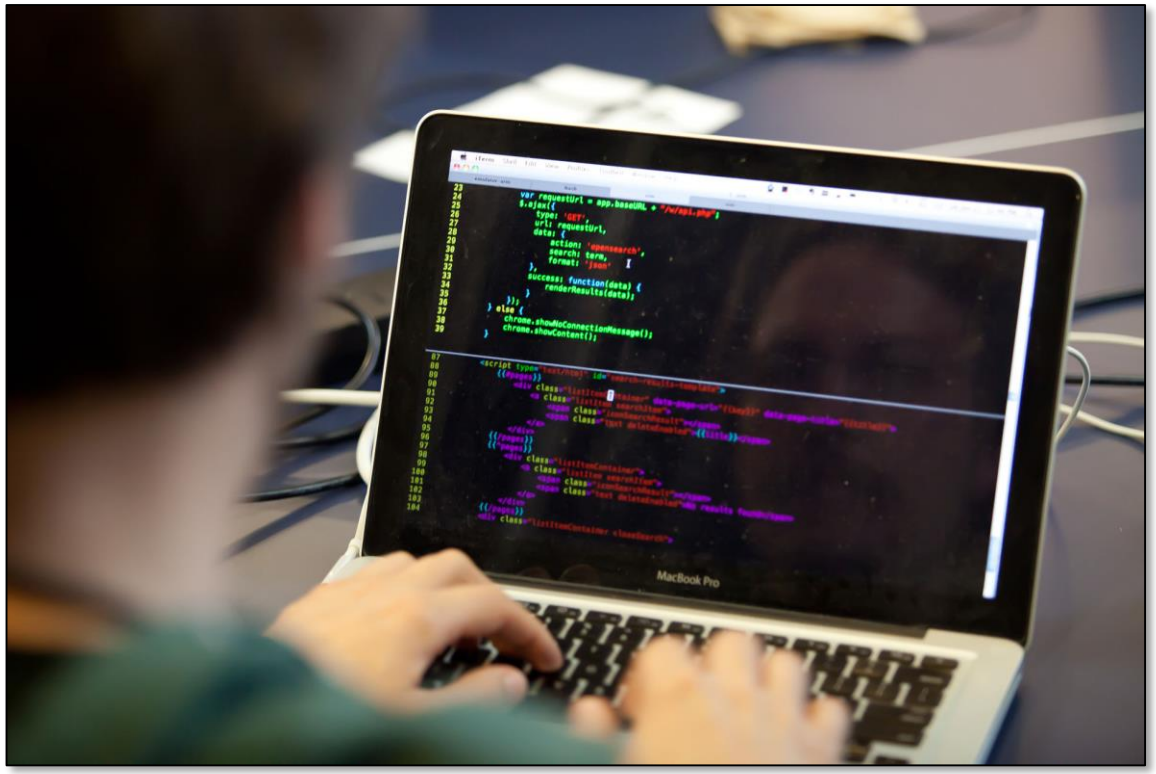
SEVENTH FRAMEWORK PROGRAMME

EU-funded project

Resources

Software availability

SAMβADA is an open source software written in C++ available at http://lasig.epfl.ch/sambada (under the license GNU GPL 3). Compiled versions are provided for Windows, Linux and MacOS X.



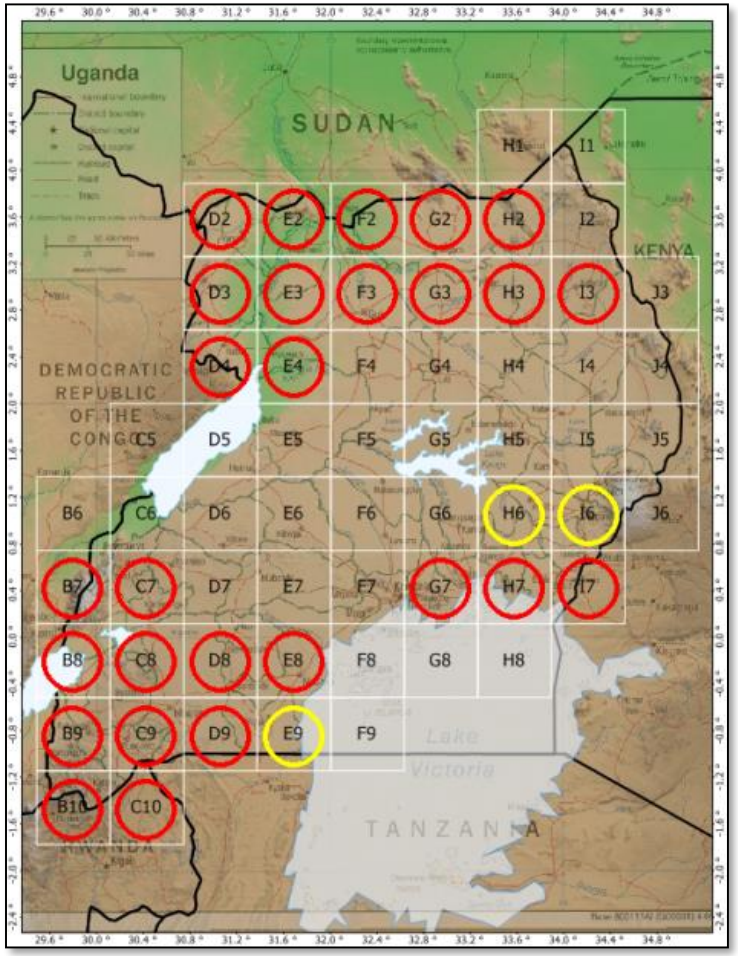
Utilization of the Scythe C++ open source library for statistical geocomputation in livestock landscape genomics

Authors

- **Sylvie Stucki**, Ecole polytechnique fédérale de Lausanne, *Switzerland*

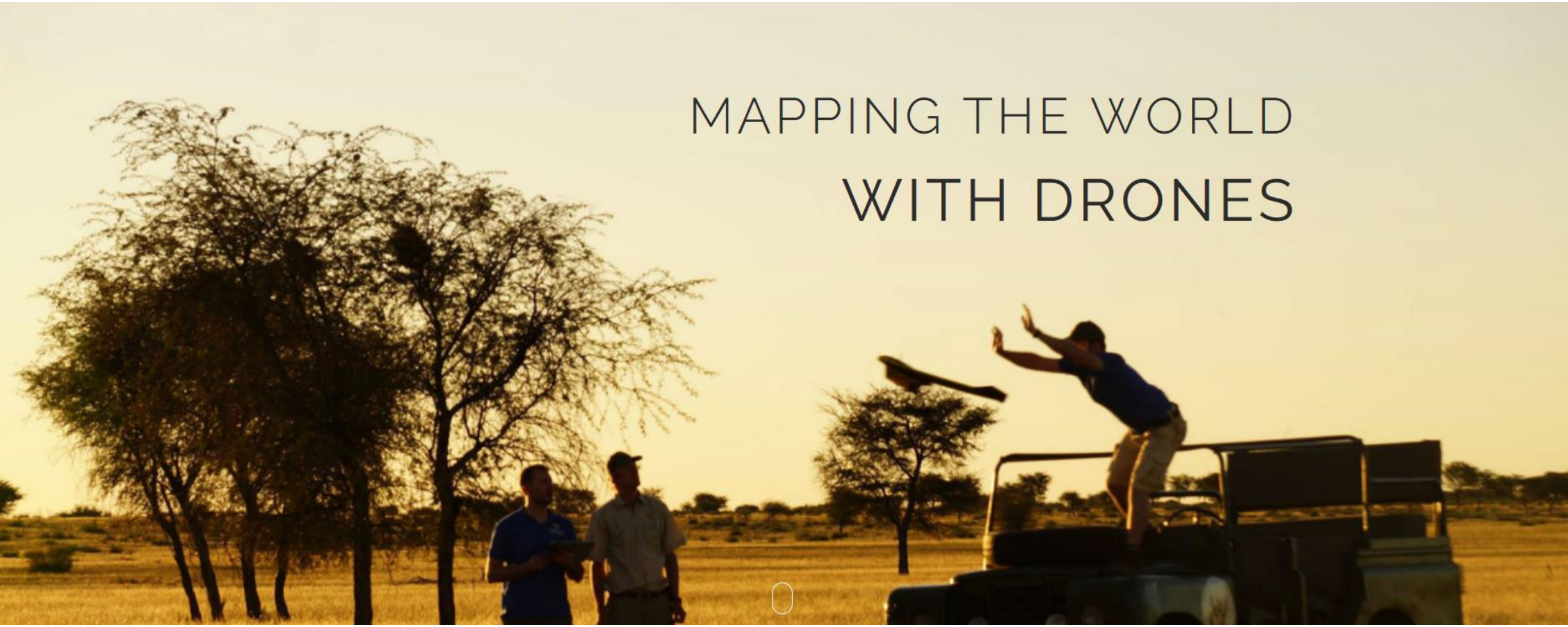
Research – open data

- Data released in accordance with the Fort Lauderdale Principles (2003)
- Public declaration in biomedicine supporting free and unrestricted use of genome sequencing data
- Data are made available under the "**Responsible use**"
- Balance between interests of the scientific community to access the data and the needs of data producers to receive recognition for their work



Research – open data

Namibia



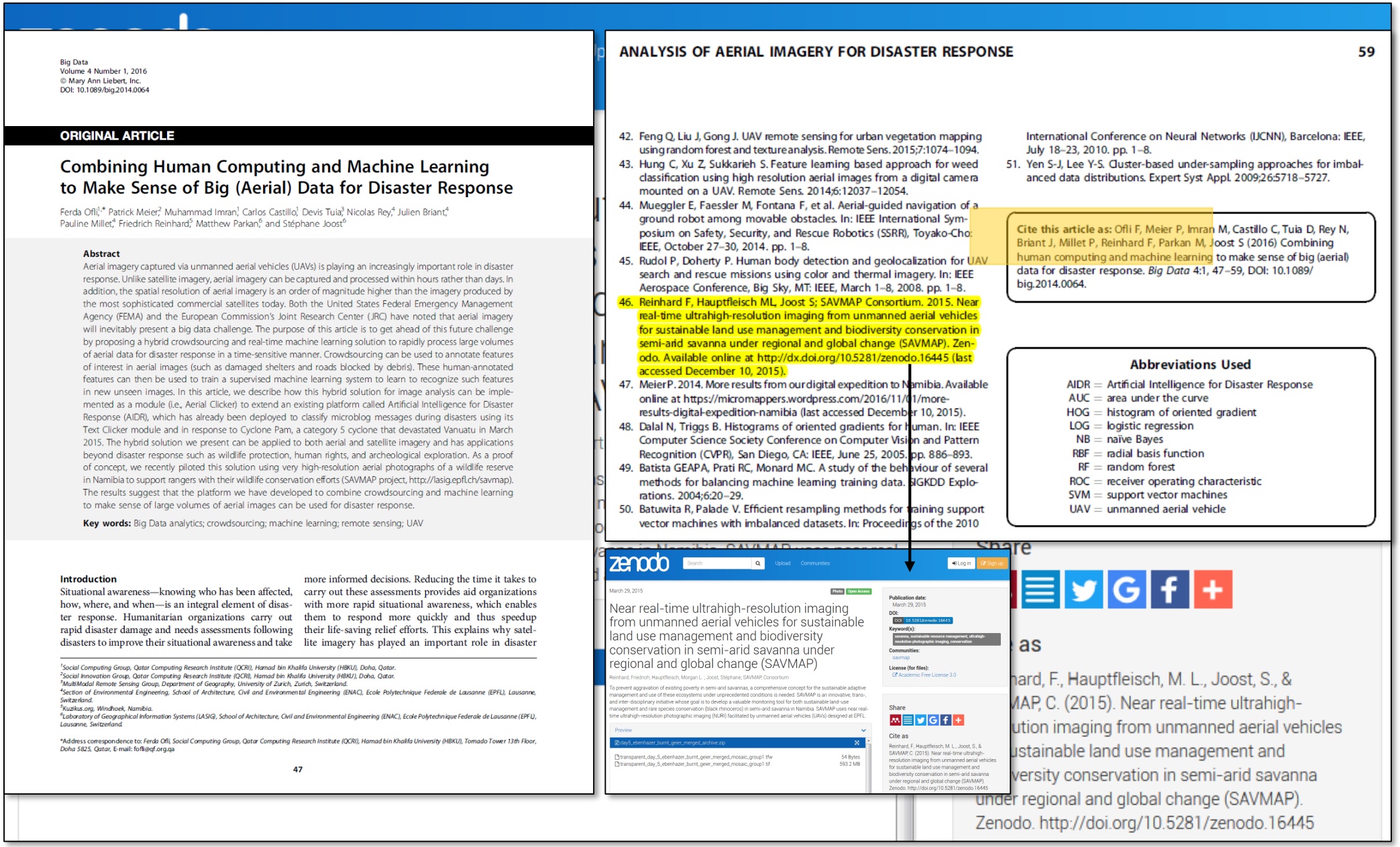
Research - Conservation of endangered species & animal counting & Vegetation survey

SAVMAP

kalahari mapping project

Research – open data

- May 2014, May 2015
- 14’000 RGB images
- 6’000 NIR images
- Next mission October 2017
- How to store and make these data open access…
- …using geographic characteristics (enabling spatial queries)
- Digital geo-repositories
- Open Data Hackdays, EPFL, November 2017



Big Data
Volume 4 Number 1, 2016
© Mary Ann Liebert, Inc.
DOI: 10.1089/big.2014.0064

ORIGINAL ARTICLE

Combining Human Computing and Machine Learning to Make Sense of Big (Aerial) Data for Disaster Response

Ferda Ofli,[1,*] Patrick Meier,[2] Muhammad Imran,[1] Carlos Castillo,[1] Devis Tuia,[3] Nicolas Rey,[4] Julien Briant,[4] Pauline Millet,[4] Friedrich Reinhard,[5] Matthew Parkan,[6] and Stéphane Joost[6]

Abstract

Aerial imagery captured via unmanned aerial vehicles (UAVs) is playing an increasingly important role in disaster response. Unlike satellite imagery, aerial imagery can be captured and processed within hours rather than days. In addition, the spatial resolution of aerial imagery is an order of magnitude higher than the imagery produced by the most sophisticated commercial satellites today. Both the United States Federal Emergency Management Agency (FEMA) and the European Commission's Joint Research Center (JRC) have noted that aerial imagery will inevitably present a big data challenge. The purpose of this article is to get ahead of this future challenge by proposing a hybrid crowdsourcing and real-time machine learning solution to rapidly process large volumes of aerial data for disaster response in a time-sensitive manner. Crowdsourcing can be used to annotate features of interest in aerial images (such as damaged shelters and roads blocked by debris). These human-annotated features can then be used to train a supervised machine learning system to learn to recognize such features in new unseen images. In this article, we describe how this hybrid solution for image analysis can be implemented as a module (i.e., Aerial Clicker) to extend an existing platform called Artificial Intelligence for Disaster Response (AIDR), which has already been deployed to classify microblog messages during disasters using its Text Clicker module and in response to Cyclone Pam, a category 5 cyclone that devastated Vanuatu in March 2015. The hybrid solution we present can be applied to both aerial and satellite imagery and has applications beyond disaster response such as wildlife protection, human rights, and archeological exploration. As a proof of concept, we recently piloted this solution using very high-resolution aerial photographs of a wildlife reserve in Namibia to support rangers with their wildlife conservation efforts (SAVMAP project, http://lasig.epfl.ch/savmap). The results suggest that the platform we have developed to combine crowdsourcing and machine learning to make sense of large volumes of aerial images can be used for disaster response.

Key words: Big Data analytics; crowdsourcing; machine learning; remote sensing; UAV

Introduction

Situational awareness—knowing who has been affected, how, where, and when—is an integral element of disaster response. Humanitarian organizations carry out rapid disaster damage and needs assessments following disasters to improve their situational awareness and take more informed decisions. Reducing the time it takes to carry out these assessments provides aid organizations with more rapid situational awareness, which enables them to respond more quickly and thus speedup their life-saving relief efforts. This explains why satellite imagery has played an important role in disaster

[1]Social Computing Group, Qatar Computing Research Institute (QCRI), Hamad bin Khalifa University (HBKU), Doha, Qatar.
[2]Social Innovation Group, Qatar Computing Research Institute (QCRI), Hamad bin Khalifa University (HBKU), Doha, Qatar.
[3]MultiModal Remote Sensing Group, Department of Geography, University of Zurich, Zurich, Switzerland.
[4]Section of Environmental Engineering, School of Architecture, Civil and Environmental Engineering (ENAC), Ecole Polytechnique Federale de Lausanne (EPFL), Lausanne, Switzerland.
[5]Kuzikus.org, Windhoek, Namibia.
[6]Laboratory of Geographical Information Systems (LASIG), School of Architecture, Civil and Environmental Engineering (ENAC), Ecole Polytechnique Federale de Lausanne (EPFL), Lausanne, Switzerland.

*Address correspondence to: Ferda Ofli, Social Computing Group, Qatar Computing Research Institute (QCRI), Hamad bin Khalifa University (HBKU), Tornado Tower 13th Floor, Doha 5825, Qatar, E-mail: fofli@qf.org.qa

42. Feng Q, Liu J, Gong J. UAV remote sensing for urban vegetation mapping using random forest and texture analysis. Remote Sens. 2015;7:1074–1094.
43. Hung C, Xu Z, Sukkarieh S. Feature learning based approach for weed classification using high resolution aerial images from a digital camera mounted on a UAV. Remote Sens. 2014;6:12037–12054.
44. Mueggler E, Faessler M, Fontana F, et al. Aerial-guided navigation of a ground robot among movable obstacles. In: IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR), Toyako-Cho: IEEE, October 27–30, 2014. pp. 1–8.
45. Rudol P, Doherty P. Human body detection and geolocalization for UAV search and rescue missions using color and thermal imagery. In: IEEE Aerospace Conference, Big Sky, MT: IEEE, March 1–8, 2008. pp. 1–8.
46. Reinhard F, Hauptfleisch ML, Joost S; SAVMAP Consortium. 2015. Near real-time ultrahigh-resolution imaging from unmanned aerial vehicles for sustainable land use management and biodiversity conservation in semi-arid savanna under regional and global change (SAVMAP). Zenodo. Available online at http://dx.doi.org/10.5281/zenodo.16445 (last accessed December 10, 2015).
47. MeierP. 2014. More results from our digital expedition to Namibia. Available online at https://micromappers.wordpress.com/2016/11/01/more-results-digital-expedition-namibia (last accessed December 10, 2015).
48. Dalal N, Triggs B. Histograms of oriented gradients for human. In: IEEE Computer Science Society Conference on Computer Vision and Pattern Recognition (CVPR), San Diego, CA: IEEE, June 25, 2005. pp. 886–893.
49. Batista GEAPA, Prati RC, Monard MC. A study of the behaviour of several methods for balancing machine learning training data. SIGKDD Explorations. 2004;6:20–29.
50. Batuwita R, Palade V. Efficient resampling methods for training support vector machines with imbalanced datasets. In: Proceedings of the 2010 International Conference on Neural Networks (IJCNN), Barcelona: IEEE, July 18–23, 2010. pp. 1–8.
51. Yen S-J, Lee Y-S. Cluster-based under-sampling approaches for imbalanced data distributions. Expert Syst Appl. 2009;26:5718–5727.

Cite this article as: Ofli F, Meier P, Imran M, Castillo C, Tuia D, Rey N, Briant J, Millet P, Reinhard F, Parkan M, Joost S (2016) Combining human computing and machine learning to make sense of big (aerial) data for disaster response. *Big Data* 4:1, 47–59, DOI: 10.1089/big.2014.0064.

Abbreviations Used

AIDR = Artificial Intelligence for Disaster Response
AUC = area under the curve
HOG = histogram of oriented gradient
LOG = logistic regression
NB = naïve Bayes
RBF = radial basis function
RF = random forest
ROC = receiver operating characteristic
SVM = support vector machines
UAV = unmanned aerial vehicle

zenodo

Near real-time ultrahigh-resolution imaging from unmanned aerial vehicles for sustainable land use management and biodiversity conservation in semi-arid savanna under regional and global change (SAVMAP)

Share

Cite as

Reinhard, F., Hauptfleisch, M. L., Joost, S., & SAVMAP C. (2015). Near real-time ultrahigh-resolution imaging from unmanned aerial vehicles for sustainable land use management and biodiversity conservation in semi-arid savanna under regional and global change (SAVMAP). Zenodo. http://doi.org/10.5281/zenodo.16445

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Knowledge must be accessible to African wildlife conservation community

LASIG - Active to spread open geoscience practices since 2008

Open source GIS tools

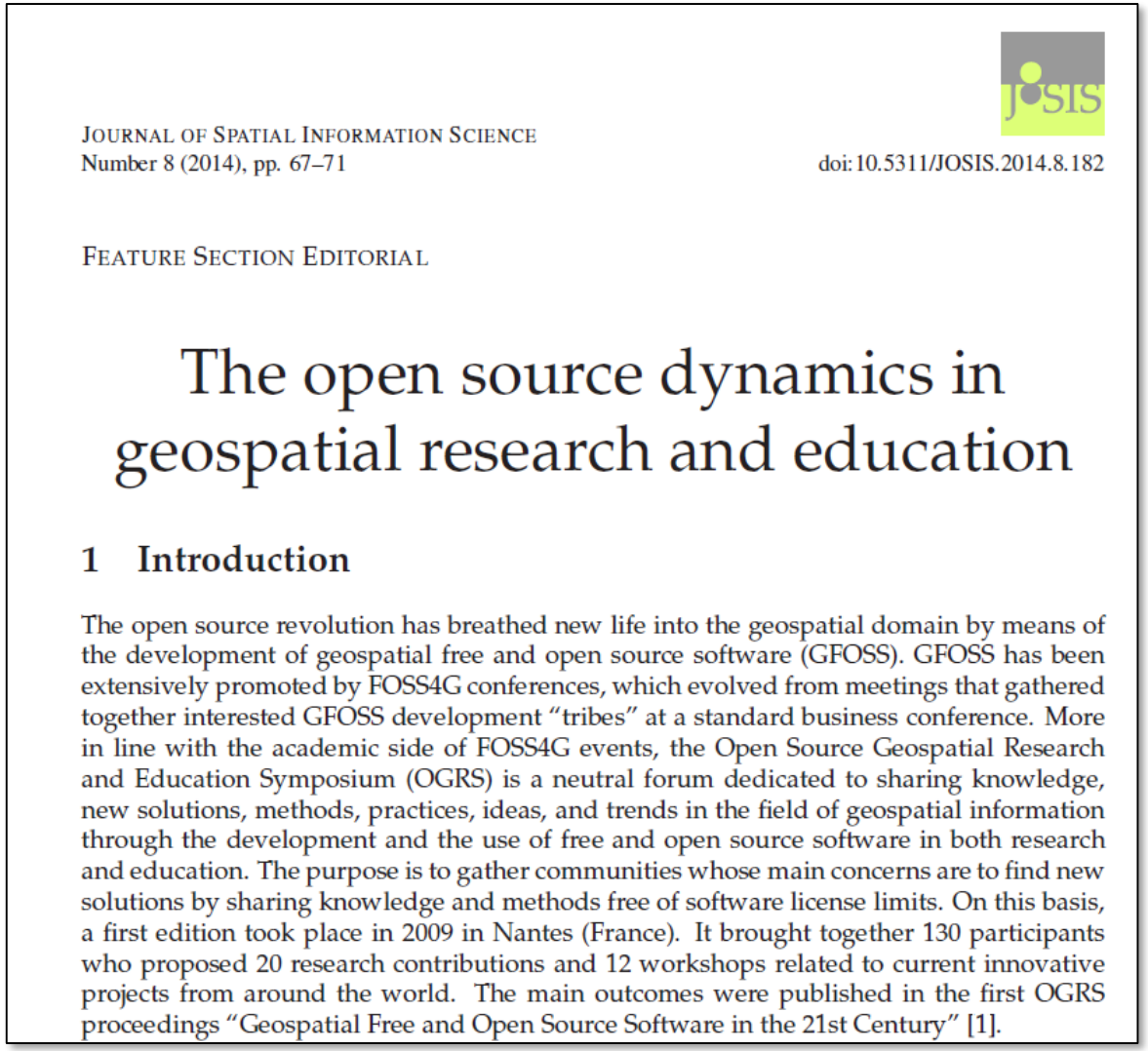
- 1978 MOSS a pioneer vector-based GIS by the US Dpt of Interior
- 1982 GRASS GIS by US Army Corps of Engineers
- 2002 QGIS – most used

Open Spatial Communities

- Free and Open Source Software for Geoinformatics 2004 (FOSS4G)
- FOSS4G 2006 Lausanne
- Quickly like any Business conference
- Education, academic
- Main outputs

http://www.ogrs-community.org/

Papers about open geospatial R&E in open access geospatial journals…



JOURNAL OF SPATIAL INFORMATION SCIENCE
Number 8 (2014), pp. 67–71

doi:10.5311/JOSIS.2014.8.182

FEATURE SECTION EDITORIAL

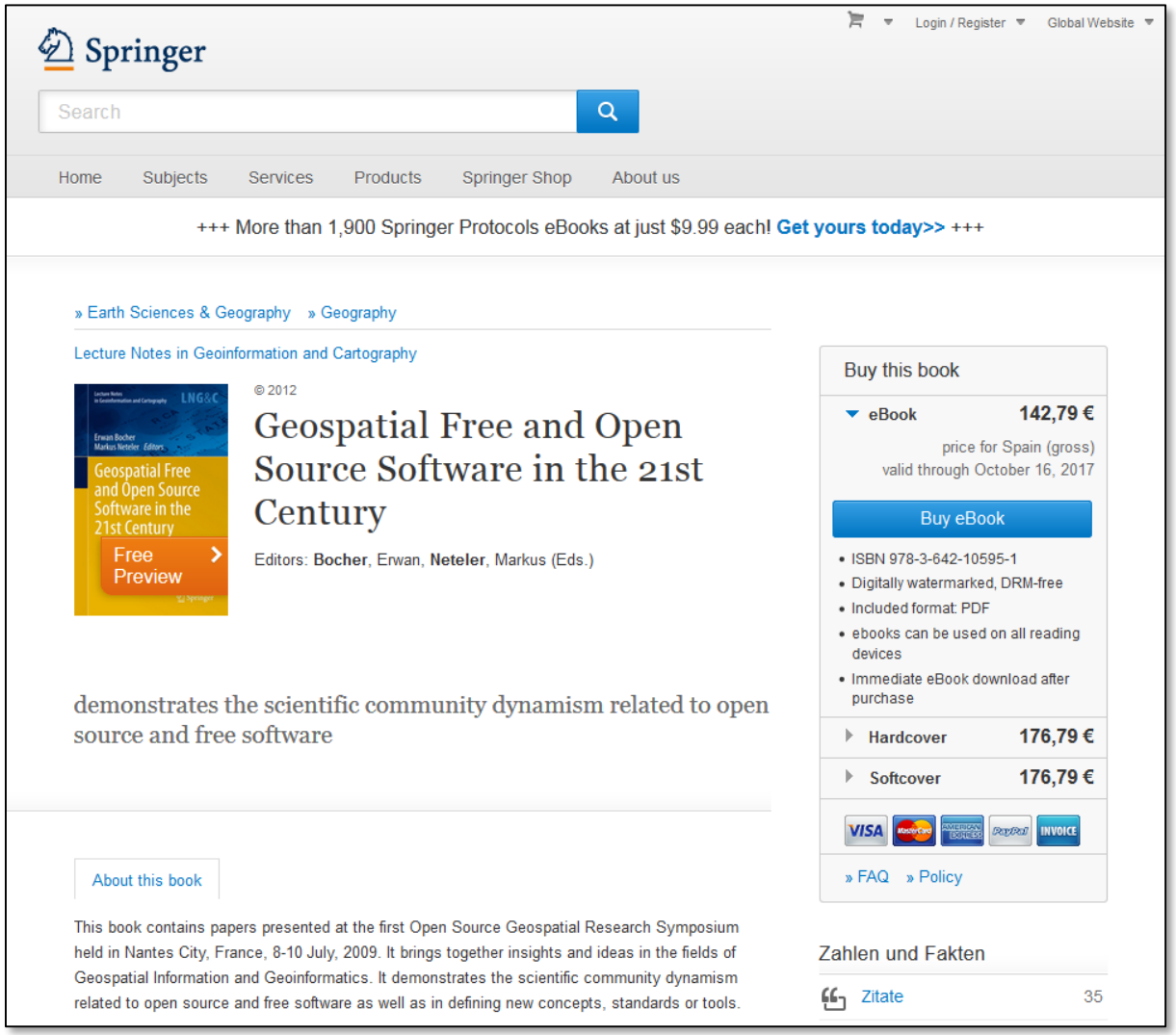
The open source dynamics in geospatial research and education

1 Introduction

The open source revolution has breathed new life into the geospatial domain by means of the development of geospatial free and open source software (GFOSS). GFOSS has been extensively promoted by FOSS4G conferences, which evolved from meetings that gathered together interested GFOSS development "tribes" at a standard business conference. More in line with the academic side of FOSS4G events, the Open Source Geospatial Research and Education Symposium (OGRS) is a neutral forum dedicated to sharing knowledge, new solutions, methods, practices, ideas, and trends in the field of geospatial information through the development and the use of free and open source software in both research and education. The purpose is to gather communities whose main concerns are to find new solutions by sharing knowledge and methods free of software license limits. On this basis, a first edition took place in 2009 in Nantes (France). It brought together 130 participants who proposed 20 research contributions and 12 workshops related to current innovative projects from around the world. The main outcomes were published in the first OGRS proceedings "Geospatial Free and Open Source Software in the 21st Century" [1].

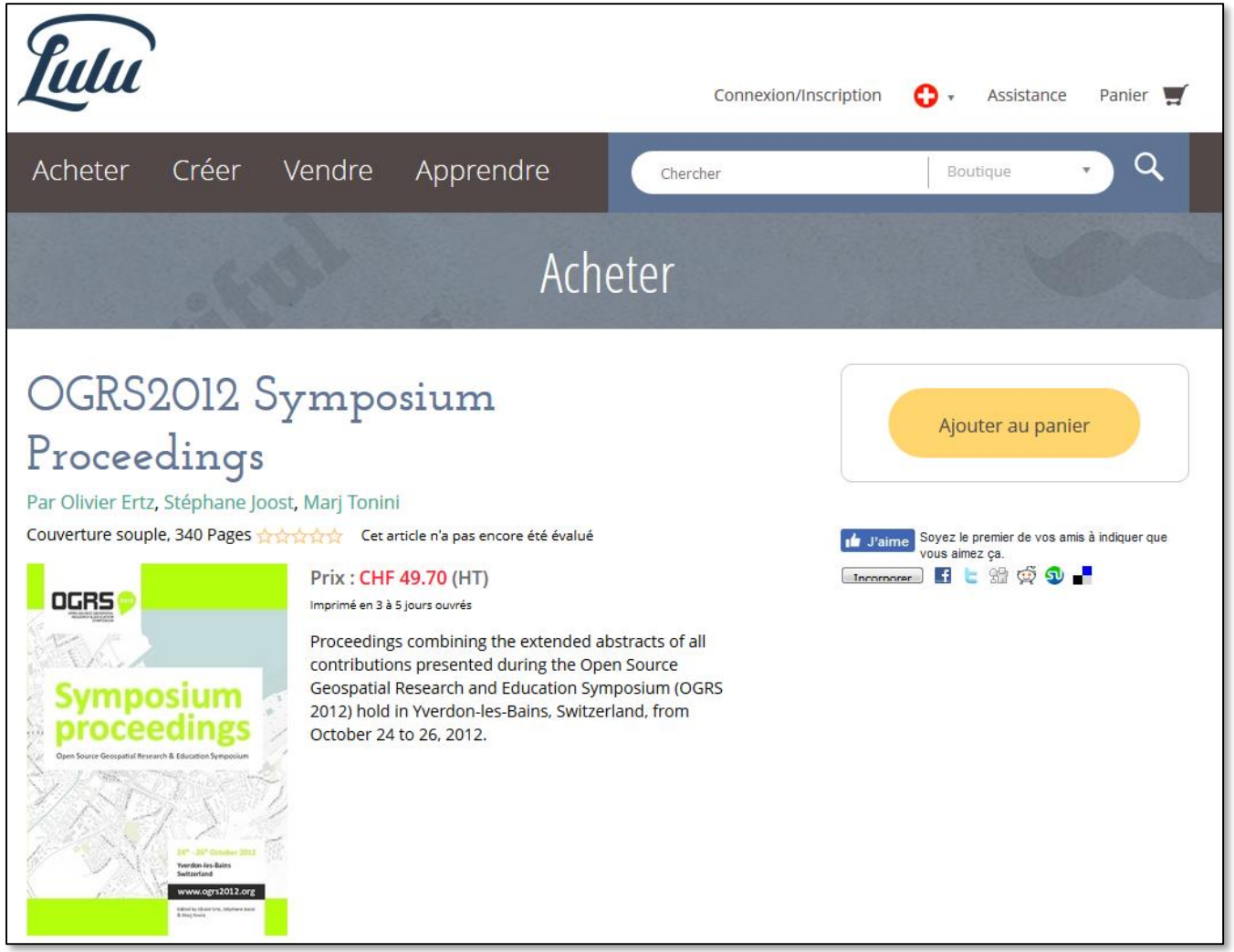
OGRS 2009, Vannes, France

- Wrong way!



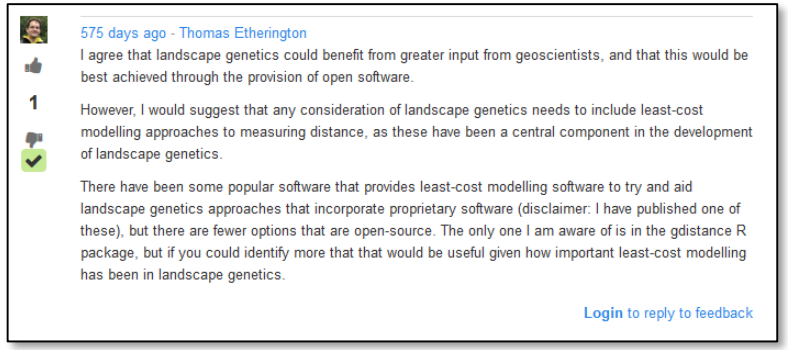
OGRS 2012, Yverdon-les-Bains, Switzerland

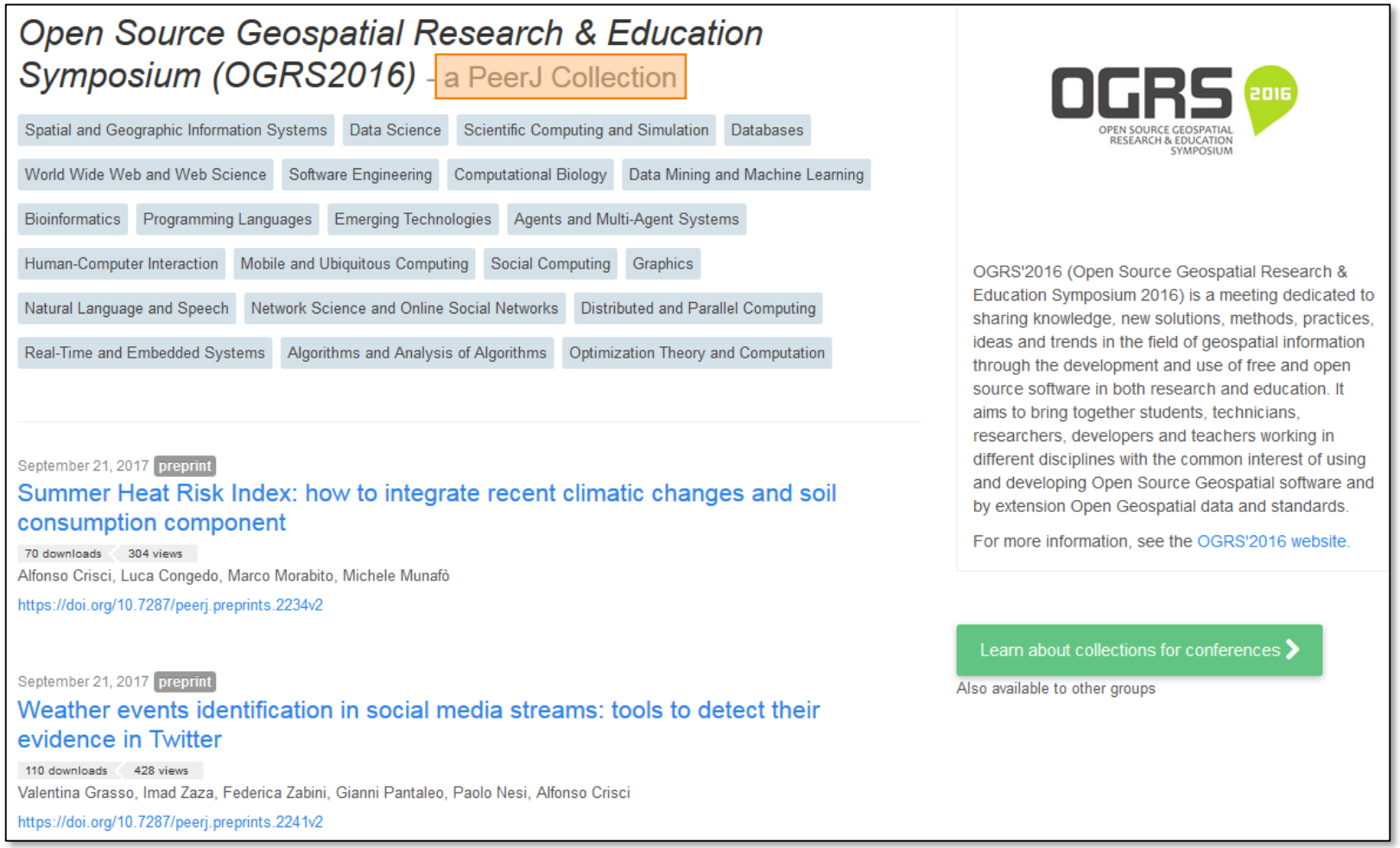
- Self-edition, SNF funds
- Dissemination through Lulu.com
- Free download, cost for the book + shipping



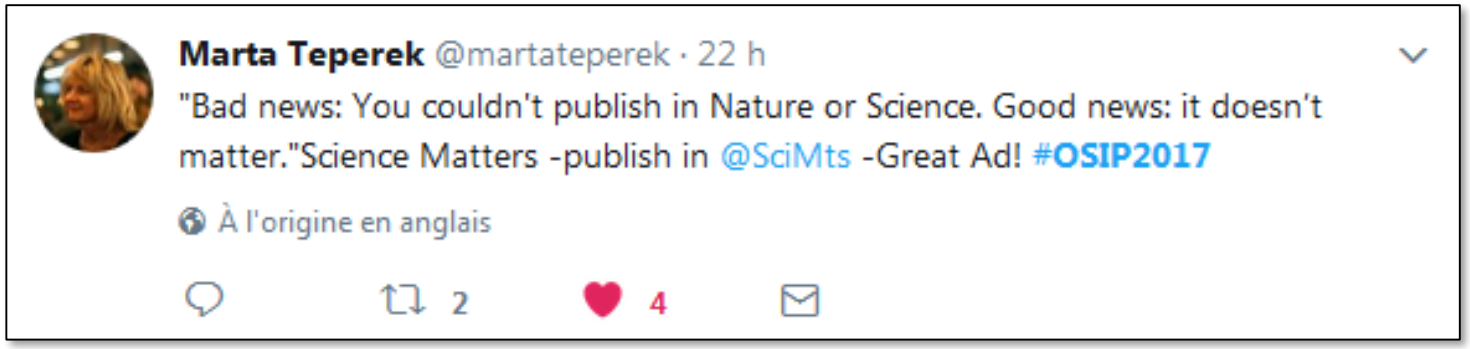
OGRS 2016, Perugia, Italy

- PeerJ platform
- Collection
- Open peer review





Conclusion



- It does not matter: it is definitely true for science
- For the while, it is not true for all scientists, at least for those who have not published in such high IF journals before…
- Looking forward to seeing how Openness will concretely be taken into account to promote the career of «open scientists»…

"… the scientific content of a paper is much more important than publication metrics or the name of the journal in which it was published", SNF

Thank you for your attention !