



ZIVILGESELLSCHAFTLICHES ENGAGEMENT

MIT

DATA SCIENCE




DATA SCIENCE

FOR

SOCIAL GOOD

Diese Folien sollen NGOs erklären, was Data Science ist und wie sie davon profitieren können.



THE BEST MINDS OF MY GENERATION
ARE THINKING ABOUT HOW TO MAKE
PEOPLE CLICK ADS. THAT SUCKS.

— Jeff Hammerbacher (Mitbegründer Facebook)

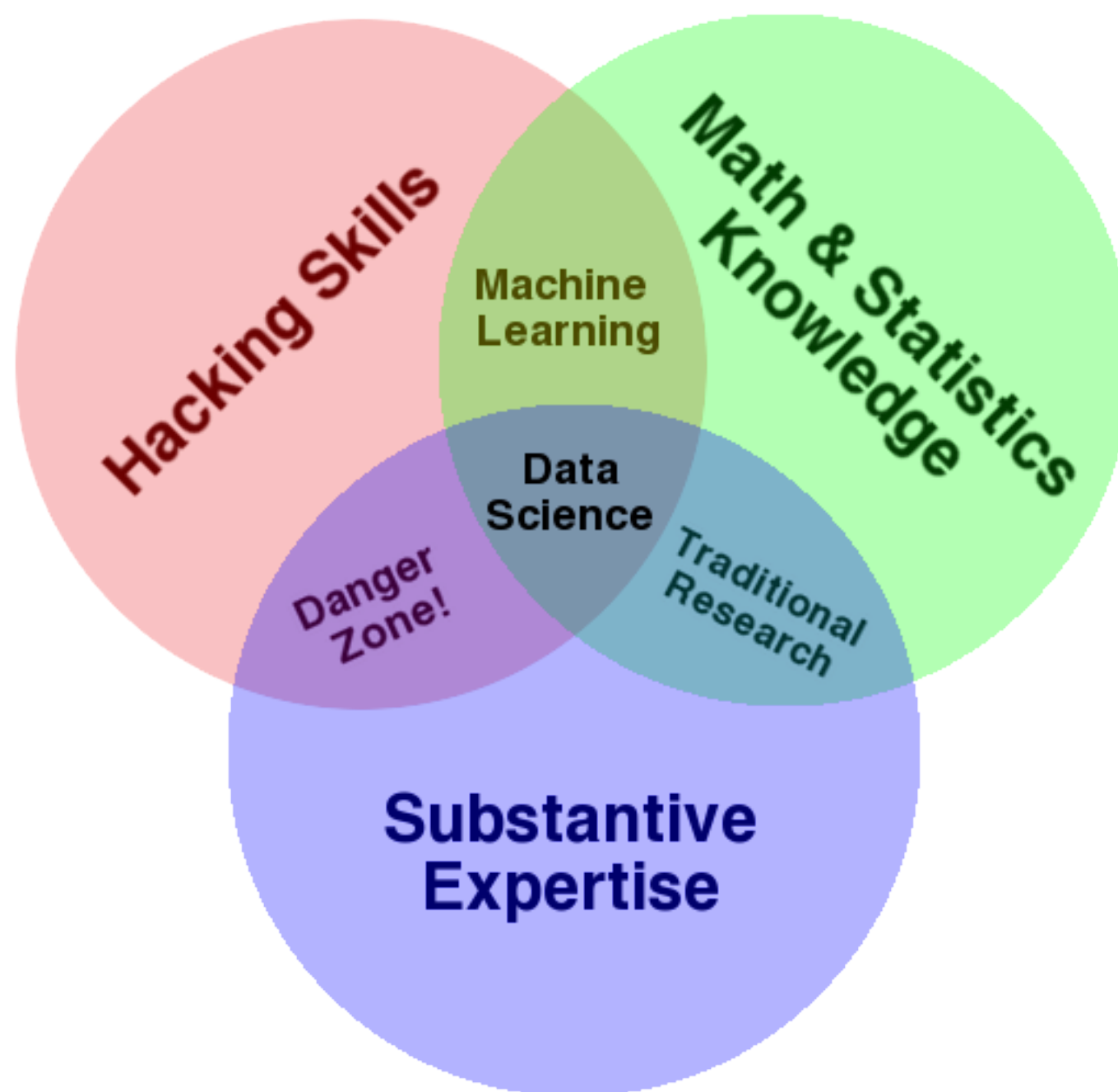
Wenn sich die Privatwirtschaft, die
Mehrheit der klugen Köpfe angelt - wer
arbeitet an den großen sozialen
Problemen unserer Zeit?

IDEE:

Zivilgesellschaftliches Engagement mit

DATA SCIENCE

WAS IST
DATA SCIENCE?



DATA SCIENCE

ist das **Ableiten von Erkenntnissen** aus **Daten**. [...]

Es werden Techniken und Theorien aus vielen Bereichen der **Mathematik, Statistik, Informatik, Maschinellem Lernen**, [...] verwendet, aus denen praxisnahe und benutzerfreundliche Vorhersagemodelle entwickelt werden. [...]

— http://en.wikipedia.org/wiki/Data_science

Das klingt etwas abstrakt... zum Glück gibt es

BEISPIELE

DATAKIND

We're tackling the world's biggest problems
through data science.

— <http://www.datakind.org>

DataKind bringt Wohltätigkeitsorganisationen
auf zwei-tägigen Data Dives mit Data Scientists
zusammen, die dort **pro bono** an
Datenproblemen der Organisationen arbeiten.



Jake Porway at TEDxMontreal

BEISPIEL: ANALYSE VON ARMUT MIT SATELLITENBILDERN

**DataKind hat GiveDirectly -
einer NGO, die bedingungslos
Geld an bedürftige Haushalte
in Kenia und Uganda
überweist¹ - geholfen,
besonders arme Dörfer durch
die automatisierte Analyse
von Satellitenbildern zu
identifizieren².**

¹ http://www.ted.com/talks/joysunshouldyoudonate_differently

² <http://www.datakind.org/projects/using-the-simple-to-be-radical/>



<https://www.youtube.com/watch?v=CxENxESpBtk>



BEISPIEL: FRÜHWARNSYSTEM FÜR MENSCHENRECHTSVERLETZUNGEN

Um zu helfen, die vielen Nachrichten, die über Amnesty Internationals Urgent Action Network eingehen, zu priorisieren, haben Freiwillige von DataKind ein Vorhersagemodell erstellt, welches die Nachrichtentexte analysiert, um auf das Eskalationspotential der Situation zu schließen.³⁴

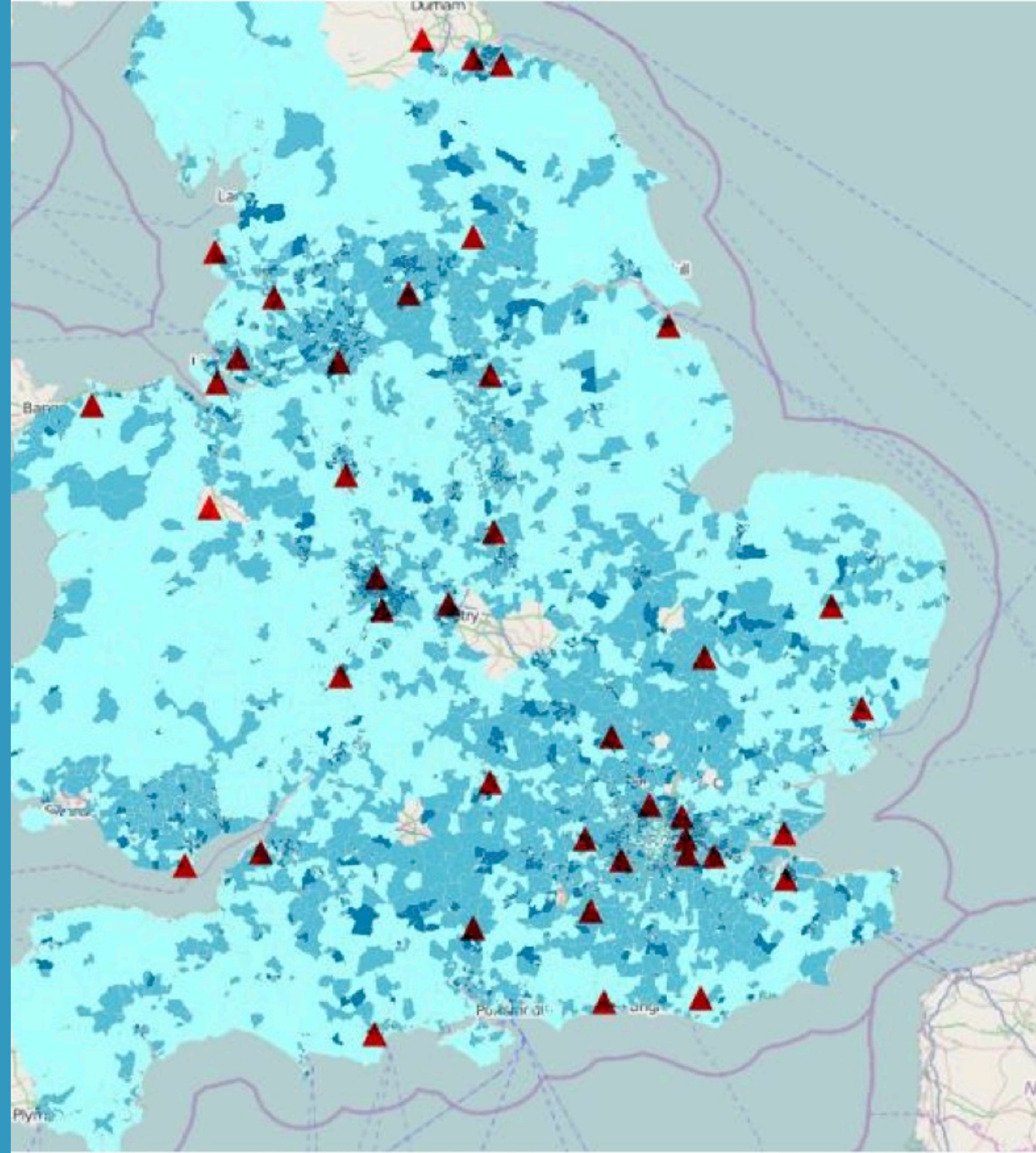
³ <http://www.datakind.org/projects/using-predictive-analytics-to-prevent-human-rights-abuses/>

⁴ http://www.washingtonpost.com/business/on-it/amnesty-international-considers-using-big-data-to-predict-human-rights-violations/2013/11/22/3f4f1a1e-5388-11e3-a7f0-b790929232e1_story.html

BEISPIEL: ANALYSE DER GEOGRAFISCHEN UNTERVERSORGUNG MIT HOSPIZEN

Freiwillige von Datakind haben Shooting Star Chase, die schwerbehinderte Kinder und ihre Familien begleiten, geholfen u.A. aktuelle und zukünftige Versorgungsengpässe in ihren Hospizen mithilfe der Vereinigung von demographischen Daten und Daten aus dem Gesundheitssektor vorherzusagen.⁵

⁵ <http://www.shootingstarchase.org.uk/our-care/info-for-families/>



Diese Projekte sind zum großen Teil von freiwilligen Data Scientists, die sonst bei Firmen wie Google, Facebook, Microsoft, uvm. arbeiten, an zwei-tägigen Data Dives umgesetzt worden.

WIE FUNKTIONIERT EIN

DATA DIVE?

VORHER: VORBEREITUNG DER DATEN

(durch freiwillige Datenbotschafter in Zusammenarbeit mit der Organisation – Vorlauf bis zu zwei Monate)

- Anonymisierung/Pseudonymisierung
 - Säuberung/Fehlerbereinigung
- Sicherstellen der Maschinenlesbarkeit der Daten-Formate

VORHER: VORBEREITUNG DER FRAGESTELLUNG

Daten offenbaren nicht auf magische Weise ihre Geheimnisse. Die richtigen Fragen müssen vorher gestellt werden, damit eine Analyse beim Data Dive Antworten liefern kann.¹⁰

¹⁰ <https://hbr.org/2013/03/you-cant-just-hack-your-way-to/>

DER DATA DIVE

Das eigentliche Event

Die teilnehmenden Organisationen „pitchen“ ihre Datenprobleme woraufhin sich die teilnehmenden freiwilligen Data Scientists um die Datenbotschafter sammeln, die der Organisation zur Seite stehen, um mithilfe von Vorhersagemodellen und Visualisierungen innerhalb eines Wochenendes Lösungen zu entwickeln.

NACHHER: IMPLEMENTIERUNG

Auch nach dem Data Dive stehen die Datenbotschafter den Organisationen zur Seite, um die prototypischen Lösungen, die auf dem Data Dive entstanden sind, in den produktiven Einsatz zu überführen.

NACHHER: IMPACT-ANALYSE

Um den Kreis zu schließen, müssen die implementierten Modelle nach einiger Zeit im Produktiveinsatz auf ihren Nutzen überprüft werden. Hat sich der Einsatz bewährt? Wo muss nachjustiert werden? Besteht weiteres/neues Potential aus den Daten zu lernen?

KLINGT TOLL! UND AUCH NOCH KOSTENLOS...
WO KANN ICH UNTERSCHREIBEN?

Noch gibt es leider keine Organisation in
Deutschland, die diese Art von
zivilgesellschaftlichem Engagement organisiert.
Das wollen **wir** ändern.

WER SIND WIR?

WIR

Wir sind Mathematiker, Statistiker, Ökonomen, Informatiker, etc. die nicht nur Werbung optimieren wollen.

- Daniel Kirsch, Diplom-Mathematiker, Berlin
 - Marit Brademann
 - Tobias Pfaff

UNSERE UNTERSTÜTZER

- Klaas Bollhöfer,
Chief Data Scientist,
The Unbelievable Machine Company
- Adam Drake,
Chief Data Scientist,
Zanox

WAS WIR BRAUCHEN

WIR BRAUCHEN PROBLEME

Aus dem Erfolg der **Data Science for Social Good**-Bewegung in den USA und England (und seit Oktober auch in Indien, Irland, und Singapur) lässt sich ableiten, dass es kein Problem sein wird, Data Scientists zu mobilisieren.

Was uns fehlt sind **NGOs**, die bessere Einsichten aus ihren **Daten** ziehen wollen.

BEI WELCHEN PROBLEMEN KÖNNEN WIR HELFEN?

Die größte Schwierigkeit für soziale Organisationen besteht darin, dass ihnen die Expertise fehlt — personell (Data Scientists) wie materiell (Big Data Technologie) — um aus ihren Daten zu lernen.

Damit einher geht die Schwierigkeit zu erkennen, welche Probleme Daten überhaupt lösen können. Wir hoffen durch die vorhergehenden Beispiele und die folgende Blaupause einen Anstoß geben zu können.

DAS TYPISCHE MUSTER

PROBLEM

Begrenzte Ressourcen müssen
möglichst **optimal** auf
Kandidaten (Personen/Gruppen/Regionen)
verteilt werden.

ANSATZ

Ein **Vorhersagemodell**
wird auf Basis von historischen Daten erstellt
um die Kandidaten zu bewerten und dadurch in
eine **Rangfolge** zu bringen.

GEWINN

Die Kandidaten,
die dem Modell nach am stärksten profitieren,
können **priorisiert** werden.

Das **spart Kosten** und maximiert den **Impact**.

HINWEIS

Auch mathematische Modelle sind nicht perfekt.

Das müssen sie aber auch nicht. Sie müssen nur besser sein, als die bisherige Praxis.

Das sind sie häufig.¹¹

¹¹ Insbesondere da sie keine Opfer von Urteilsheuristiken und kognitiven Verzerrungen sind.

ANDERE TYPISCHE ANWENDUNGSFÄLLE

- Ausreißeranalyse/Anomalieerkennung (z.B. um Missbrauch aufzudecken)
 - Social Media Analyse
 - ???

KONTAKT

Daniel Kirsch
mail@danielkirs.ch
@kirel

Webseite ist in Arbeit.

WEITERFÜHRENDE LINKS:

- <http://www.nominettrust.org.uk/knowledge-centre/articles/big-data-and-social-organisations>
- <http://datakind.org>
- <http://dssg.io>

WEITERFÜHRENDE LINKS (FORTSETZUNG):

- <http://dssg-atl.io/>
- <http://bayesimpact.org>
- <http://codefor.de>

ATtribution

- Foto von Jeff Hammbacher von Fred Brenenson lizenziert unter CC BY 2.0