Workflow & tasks



SPIEGEL ONLINE

collecting data

scraping data when released as HTML tables

turning pdf reports into well structured data (data journalism team)

processing data

merge data sets released by different pharmaceutical companies deduplication / person matching setting up data base

validation of extracted data sets (data journalism team)

extensive data base articles

compact data base interactive map articles (science / health department)

reporting

collaborative reporting

Turn nasty pdf-files into clean, well-structured data

- different layout/structures
- password protected pdf-files
- some deliberately ill-structured
- some image pdfs

As if they wanted the data to be trapped...

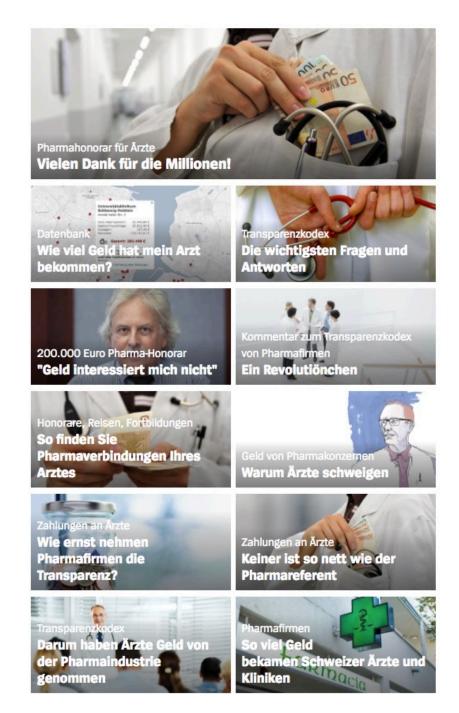
Workflow: (Remove password protection) – (OCR) – Extracing data from pdf (Tabula) - Validating/correcting extraction (Excel / R)

Want to hear & learn more about this → Stay for the workshop 15:15 – 16:15



How to present the findings?

- Overview
- Search in database + map
- FAQ
- Story / confrontation
- Opinion piece
- Bigger picture / explainer
- Individual cases



What readers are most interested in

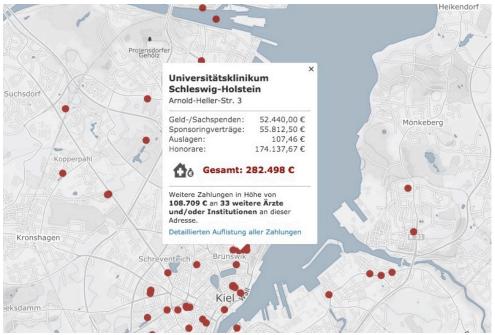
Finding their doctor(s)

→ Enormous traffic on data base and map based search



Deine Suche ergab folgende Treffer

Name	Adresse	Gesamtbetrag
med update GmbH	Hagenauer Str. 53 65203 Wiesbaden	1.453.450,00 €
Rheumatologische Fortbildungsakademie GmbH	Köpenicker Str. 48/49 10179 Berlin	1.087.391,86 €
Charité Universitätsmedizin Berlin	Charitéplatz 1 10117 Berlin	1.041.489,06 €
Medical Tribune Verlagsgesellschaft mbH	Unterden Eichen 5 65195 Wiesbaden	945.934,78 €



Outcome

Readers

High interest, mostly positive, encouraging feedback

Doctors

Mixed feedback, frustration with current transparency regulations and how education is dominated by pharmaceutical companies

Lawyers

Numerous complaints, systematic targeting of doctors by at least one legal practice. But: republication of transparency reports, that are based on individual consent is legal

Politics

no changes