



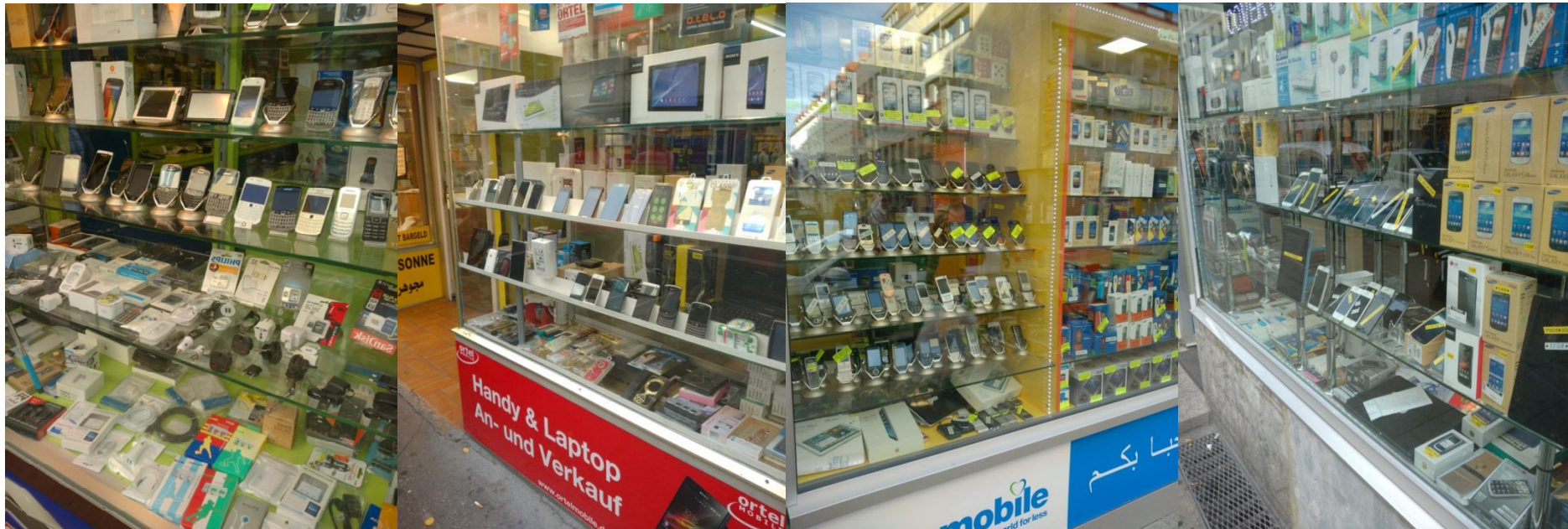
Telling the Story of People's Behavior with Smartphone Data

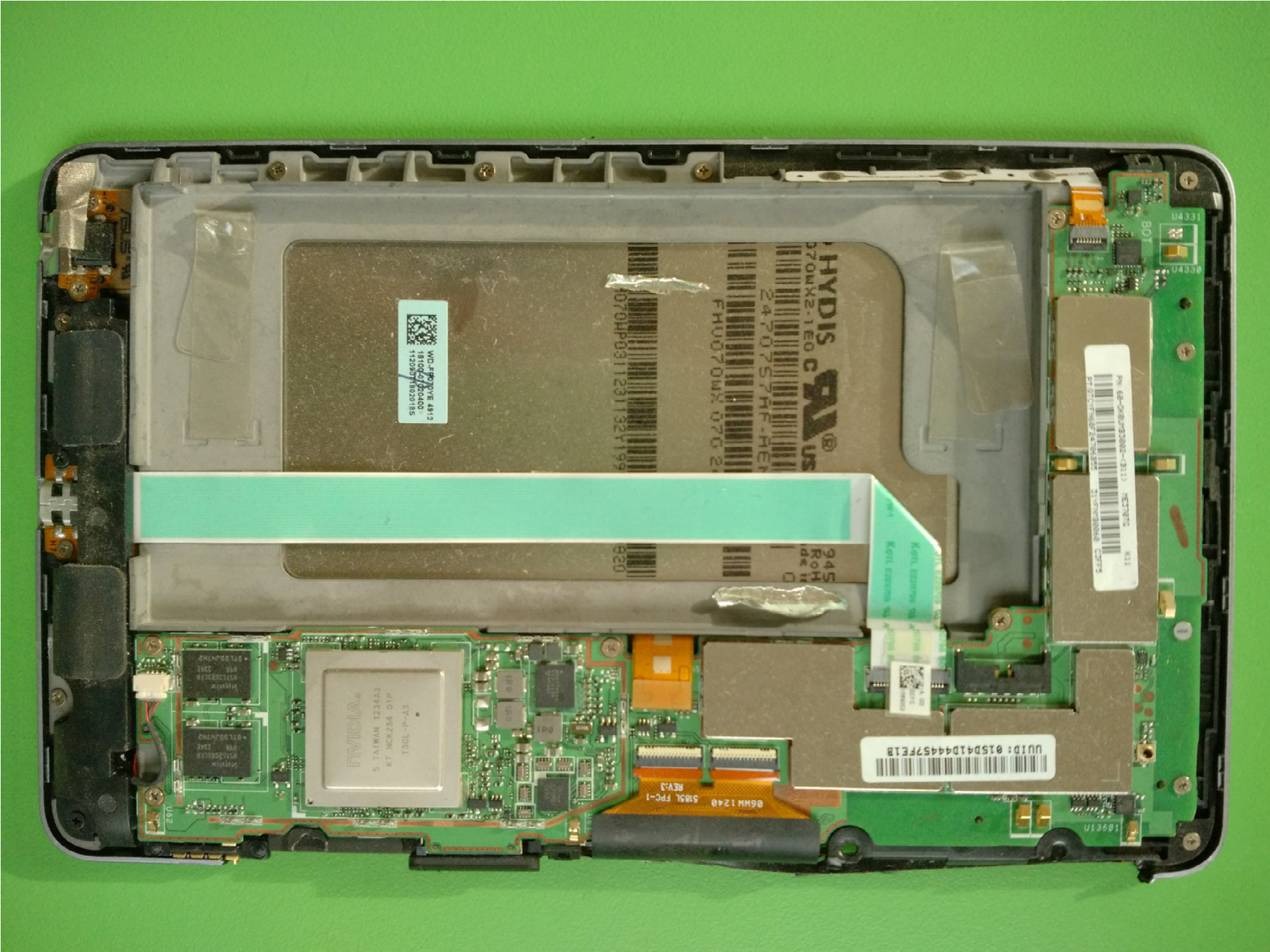
Joerg Blumtritt
@jbenno



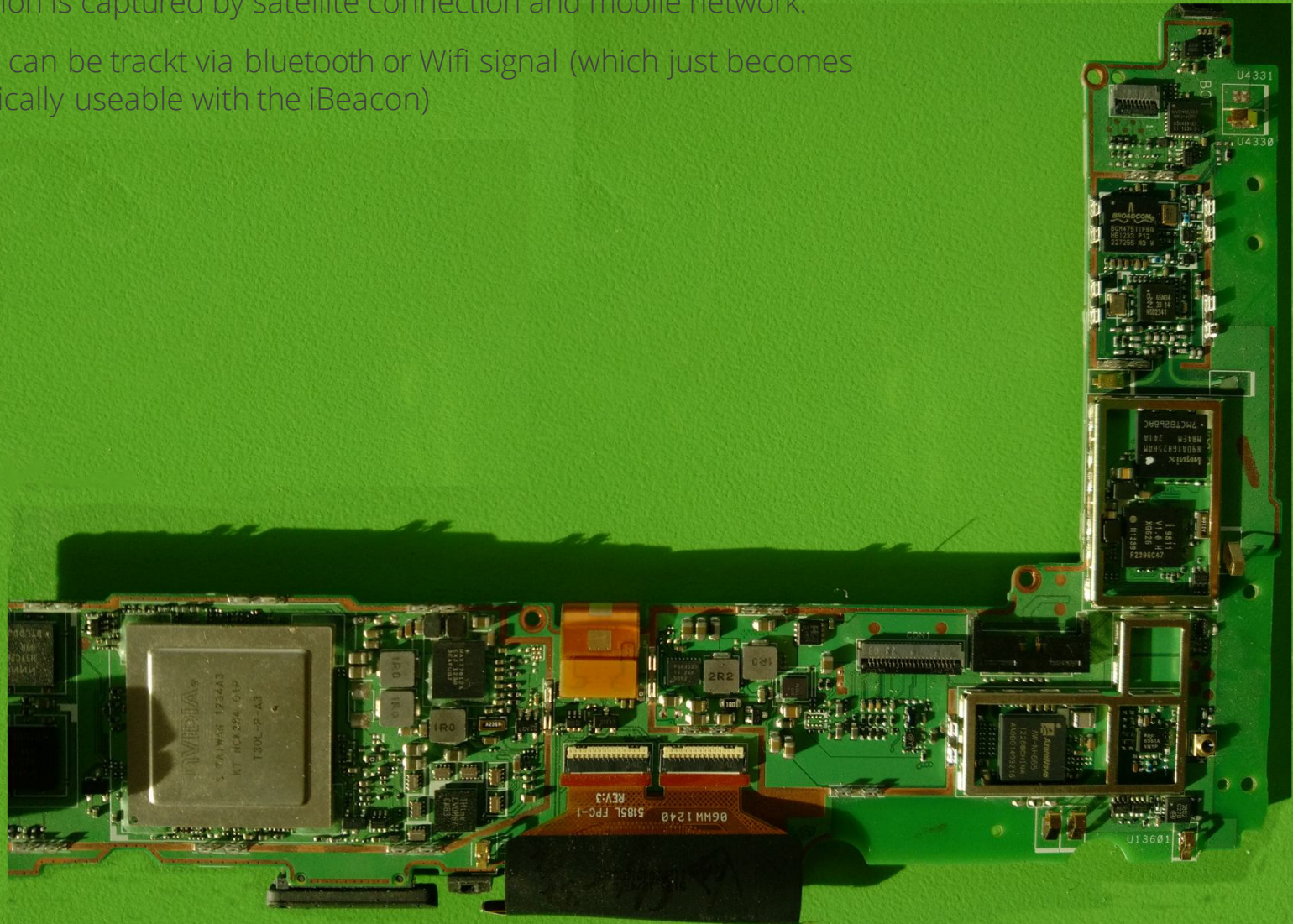
Mobile

- Two billion people use smartphones (three times more than users of PCs)
- Smartphones are far more than just „mobile computers“: they carry more than twenty sensors or probes, continuously monitoring our behavior and our environment.
- Mobile is becoming the most important source of human generated data and surpasses social networks.
- Apart from people using their phones, there are more than five billion mobile devices, connected to objects, like e.g. cars. These build the Internet of Things.





- Smartphones carry a phalanx of sensors and track all kind of environmental data.
- Our movements and immediate surroundings are monitored by gyroscope, accelerometer, luminosity sensor in the camera, microphone etc.
- The location is captured by satellite connection and mobile network.
- Proximity can be tracked via bluetooth or Wifi signal (which just becomes systematically useable with the iBeacon)



MEMS

Gyroscope,
Accelerometer

Satellites

GPS

Microphone,
Camera
Temperature,
Air Pressure,
Compass, ...

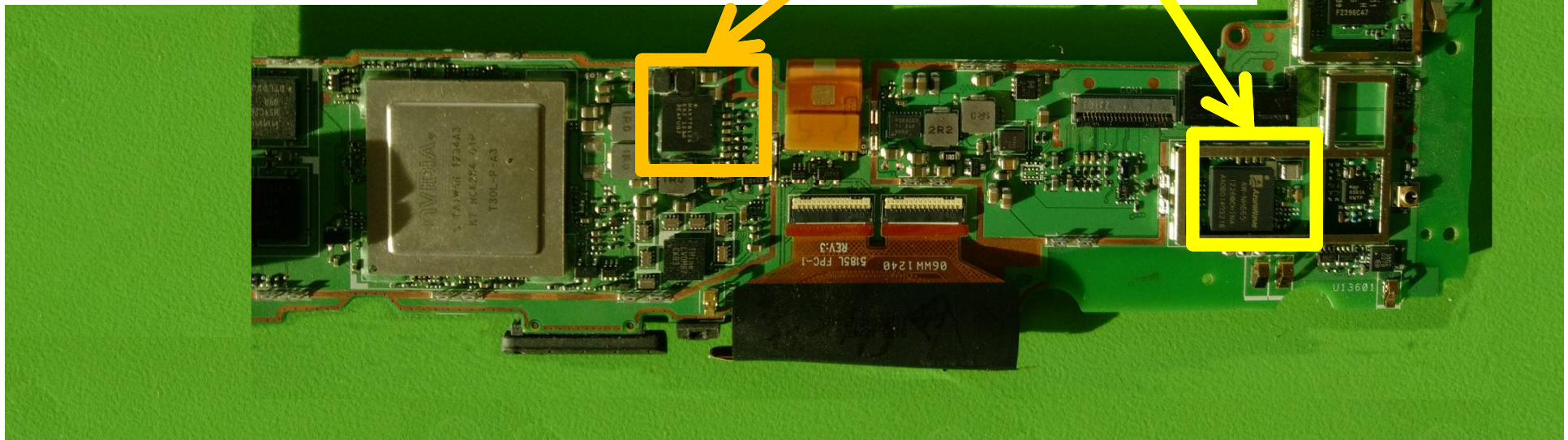
Radio

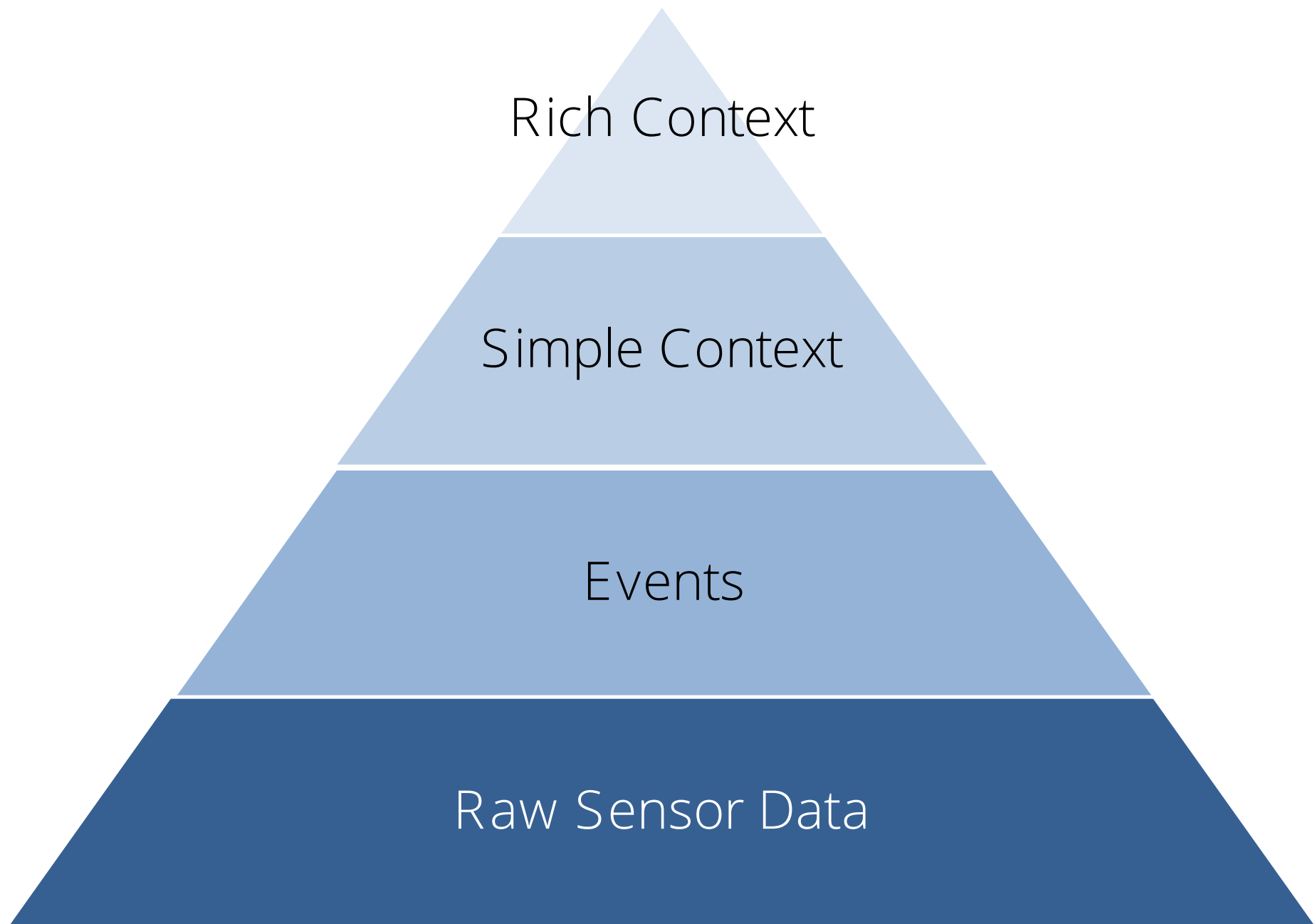
NFC

Wifi, Bluetooth
4G

Supplies

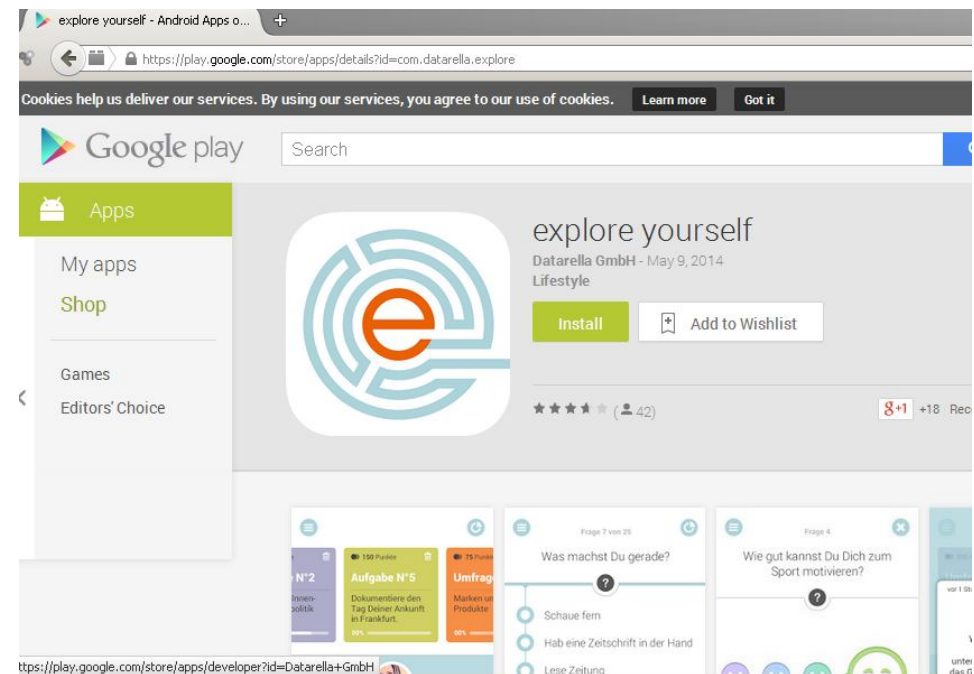
Battery

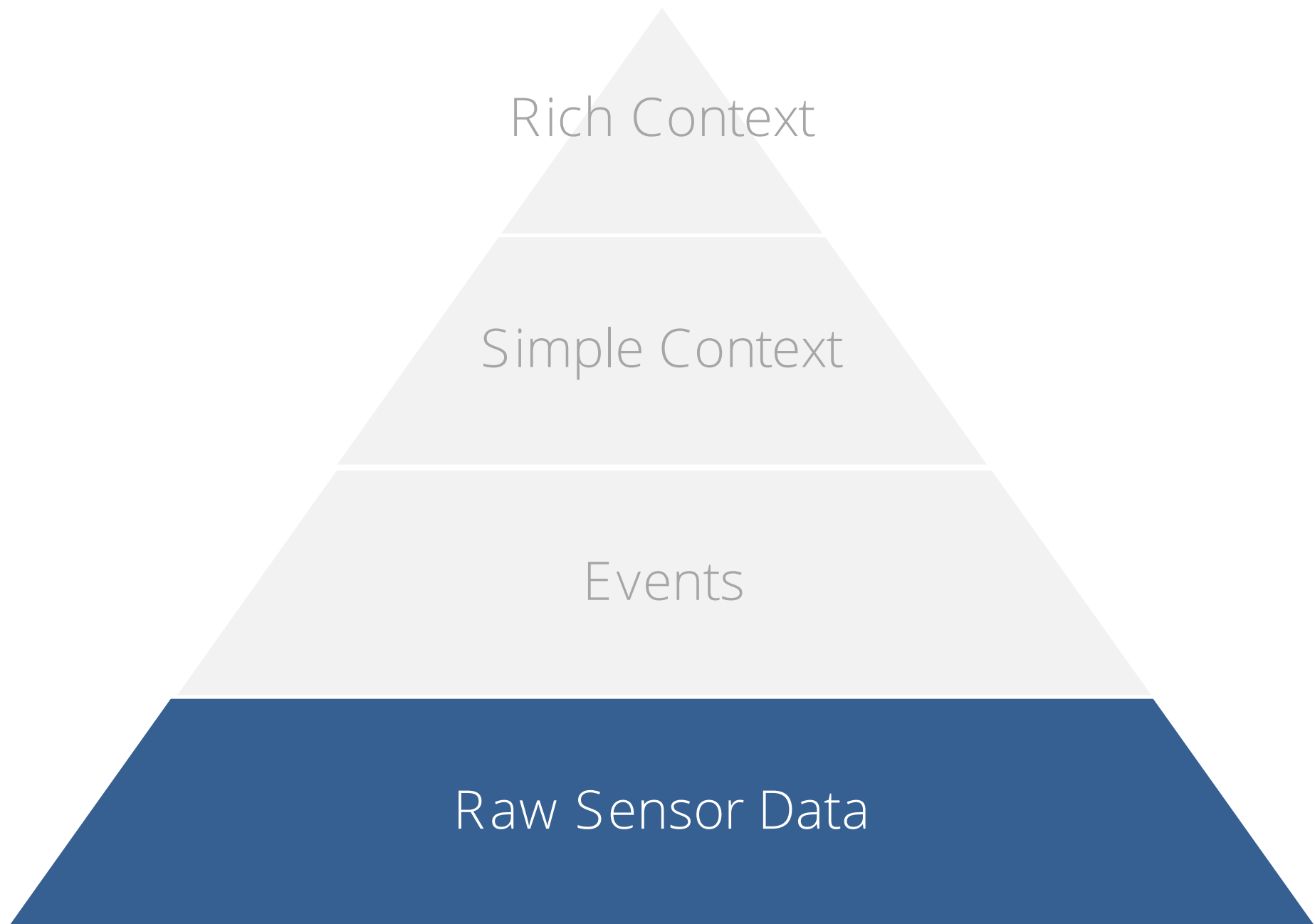




Our App: explore

- We started our own app 'explore':
- explore tracks all kinds of sensor data on the smartphone. The data can be collected for analysis, and it can trigger interactions (like asking questions or offering suggestions).
- The open beta is available on Google Play Store; the iOS version should be ready by July 2014.

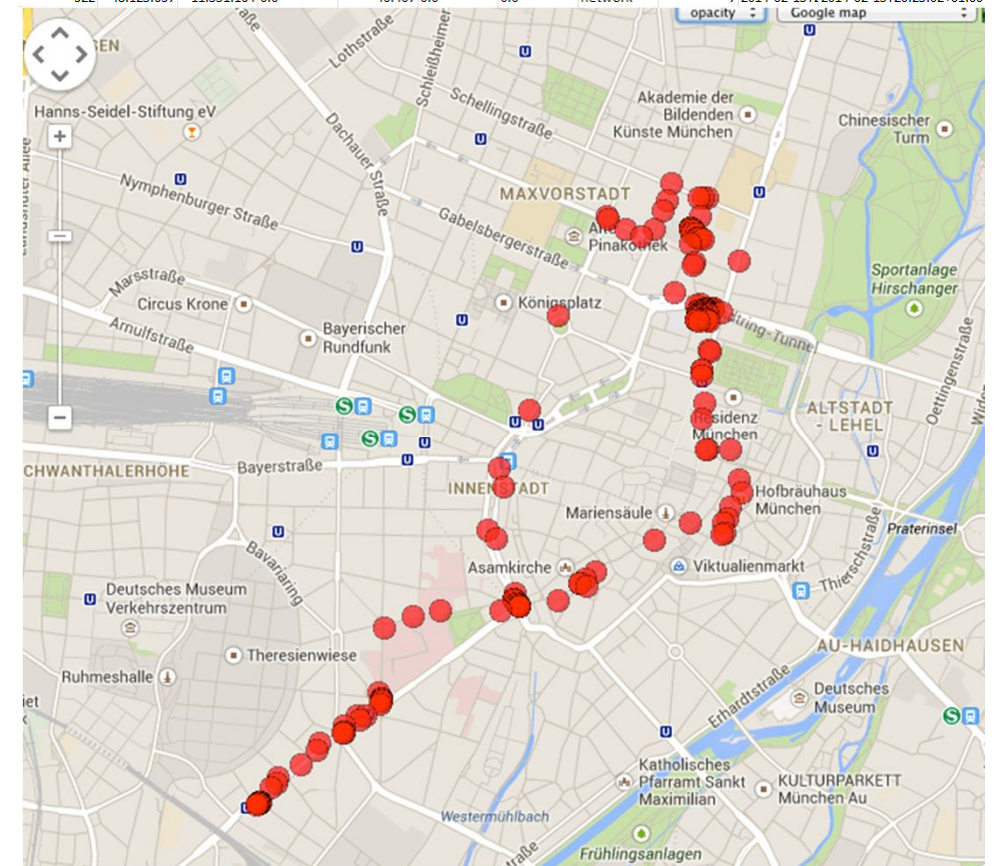




Data

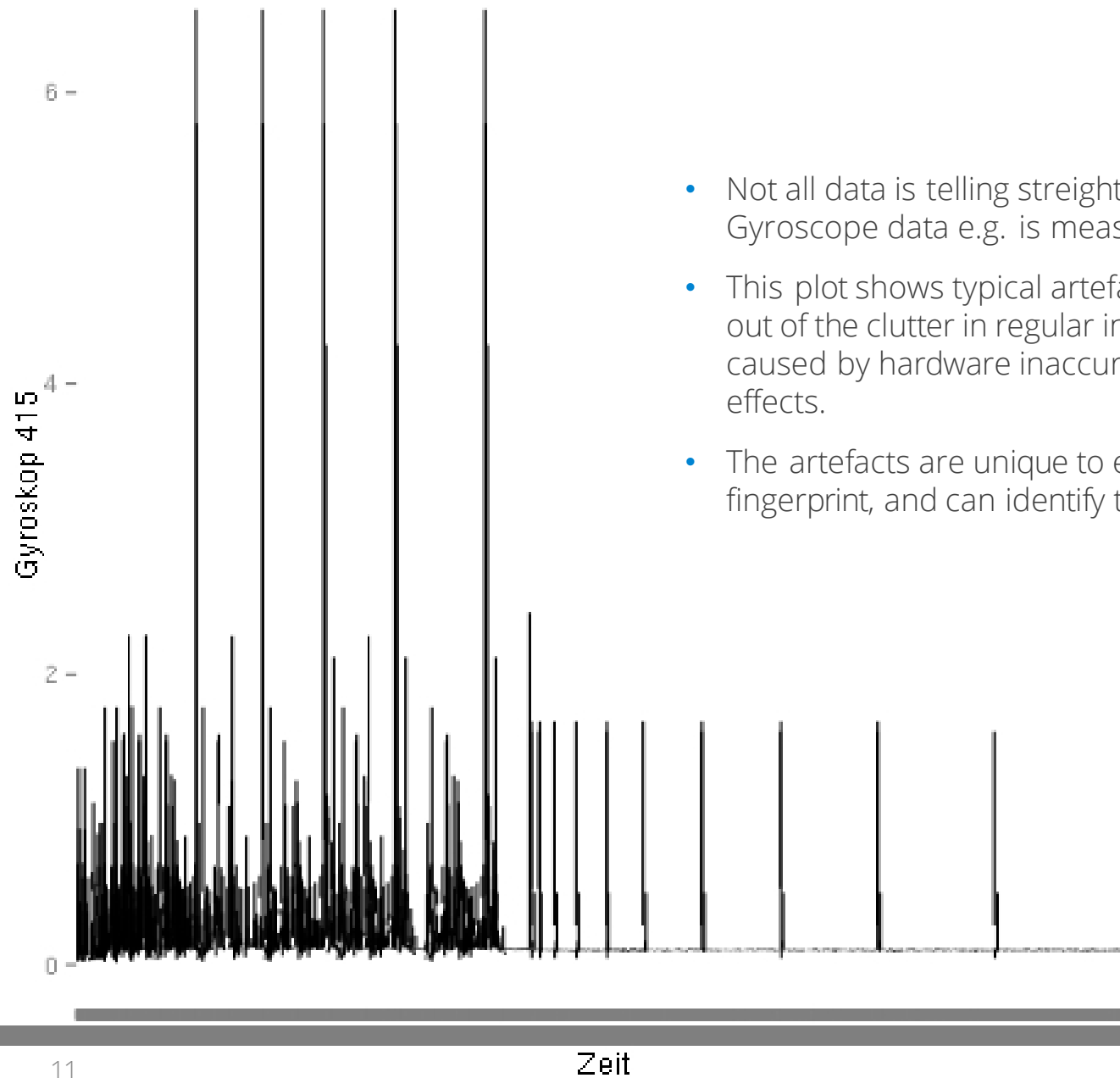
- Sensor data is generated mostly in forms of tables, locally stored as SQL databases for each app. We transfer the data to analyze it, e.g. visualize geo-location on a map.

id	latitude	longitude	altitude	accuracy	bearing	speed	provider	user_id	recorded	created_a
300	48.125.637	11.551.185	0.0	39.826	0.0	0.0	network	7	2014-02-15T12:04:02+01:00	2014-02-15T17:53:03+01:00
301	48.125.664	11.551.182	0.0	37.745	0.0	0.0	network	7	2014-02-15T12:04:02+01:00	2014-02-15T18:12:01+01:00
302	48.125.652	11.551.146	0.0	39.509	0.0	0.0	network	7	2014-02-15T12:04:02+01:00	2014-02-15T18:12:02+01:00
303	48.125.652	1.155.116	0.0	39.21	0.0	0.0	network	7	2014-02-15T12:04:02+01:00	2014-02-15T18:12:03+01:00
304	48.125.645	11.551.162	0.0	39.654	0.0	0.0	network	7	2014-02-15T12:04:02+01:00	2014-02-15T18:31:02+01:00
305	48.125.652	11.551.164	0.0	38.991	0.0	0.0	network	7	2014-02-15T12:04:02+01:00	2014-02-15T18:31:02+01:00
306	48.125.645	11.551.177	0.0	24.935	0.0	0.0	network	7	2014-02-15T12:04:02+01:00	2014-02-15T18:31:03+01:00
307	48.125.652	1.155.117	0.0	38.834	0.0	0.0	network	7	2014-02-15T12:04:02+01:00	2014-02-15T18:50:02+01:00
308	4.812.565	11.551.186	0.0	38.818	0.0	0.0	network	7	2014-02-15T12:04:02+01:00	2014-02-15T18:50:02+01:00
309	48.125.652	11.551.168	0.0	38.95	0.0	0.0	network	7	2014-02-15T12:04:02+01:00	2014-02-15T18:50:02+01:00
310	4.812.565	11.551.159	0.0	39.538	0.0	0.0	network	7	2014-02-15T12:04:02+01:00	2014-02-15T19:09:01+01:00
311	48.125.664	11.551.157	0.0	38.043	0.0	0.0	network	7	2014-02-15T12:04:02+01:00	2014-02-15T19:09:02+01:00
312	48.125.664	11.551.175	0.0	37.869	0.0	0.0	network	7	2014-02-15T12:04:02+01:00	2014-02-15T19:09:03+01:00
313	48.125.652	11.551.172	0.0	38.873	0.0	0.0	network	7	2014-02-15T12:04:02+01:00	2014-02-15T19:28:02+01:00
314	48.125.664	1.155.116	0.0	38.043	0.0	0.0	network	7	2014-02-15T12:04:02+01:00	2014-02-15T19:28:02+01:00
315	4.812.565	1.155.117	0.0	39.186	0.0	0.0	network	7	2014-02-15T12:04:02+01:00	2014-02-15T19:28:03+01:00
316	48.125.656	11.551.165	0.0	38.686	0.0	0.0	network	7	2014-02-15T12:04:02+01:00	2014-02-15T19:47:02+01:00
317	48.125.664	11.551.159	0.0	38.194	0.0	0.0	network	7	2014-02-15T12:04:02+01:00	2014-02-15T19:47:02+01:00
318	4.812.562	11.551.171	0.0	22.996	0.0	0.0	network	7	2014-02-15T12:04:02+01:00	2014-02-15T19:47:03+01:00
319	48.125.637	11.551.172	0.0	24.383	0.0	0.0	network	7	2014-02-15T12:04:02+01:00	2014-02-15T20:06:02+01:00
320	4.812.564	11.551.148	0.0	25.239	0.0	0.0	network	7	2014-02-15T12:04:02+01:00	2014-02-15T20:06:02+01:00
321	4.812.563	11.551.156	0.0	23.97	0.0	0.0	network	7	2014-02-15T12:04:02+01:00	2014-02-15T20:06:03+01:00
322	48.125.637	11.551.164	0.0	40.407	0.0	0.0	network	7	2014-02-15T12:04:02+01:00	2014-02-15T20:25:02+01:00

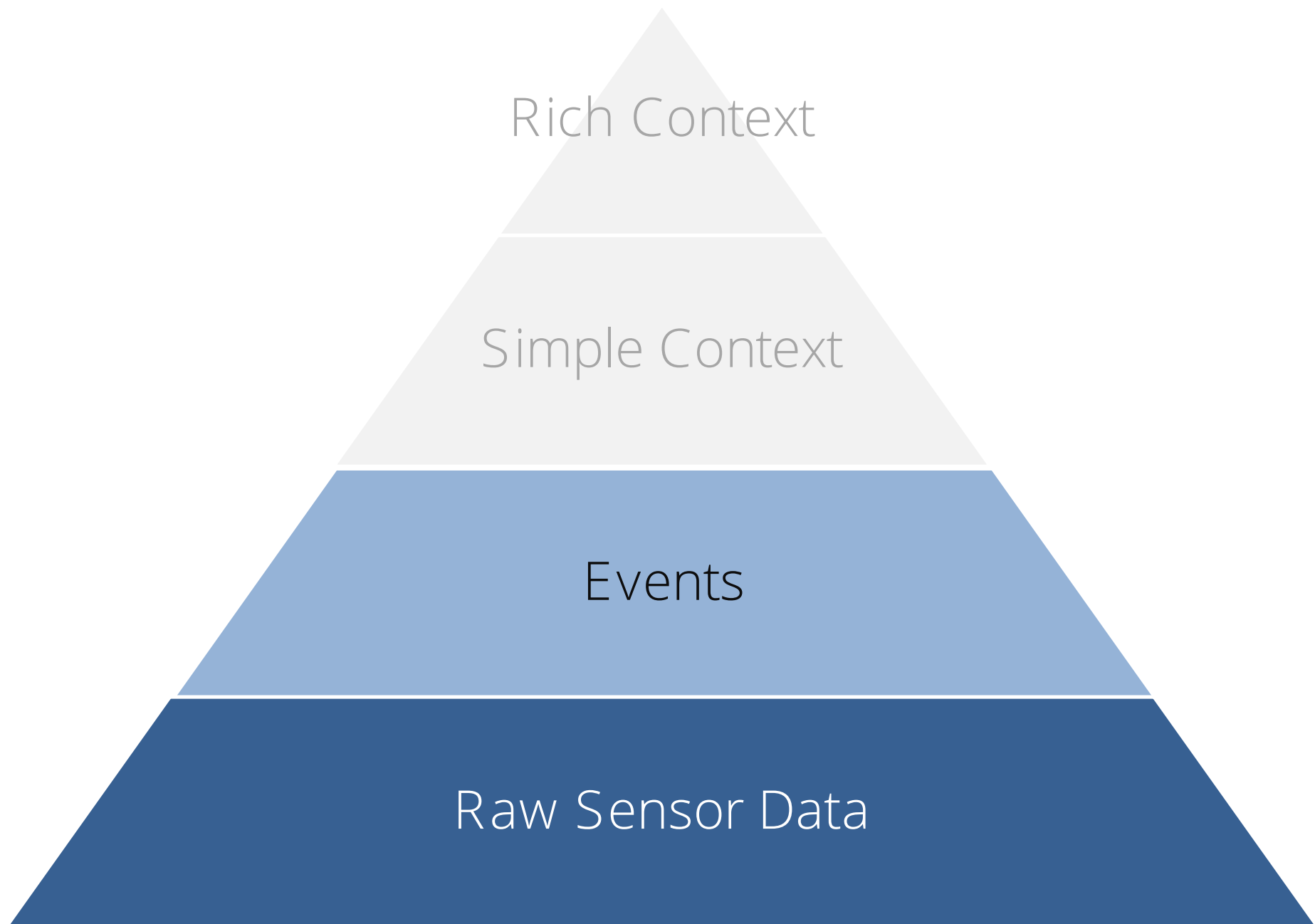


Datarella - Joerg Blumtritt



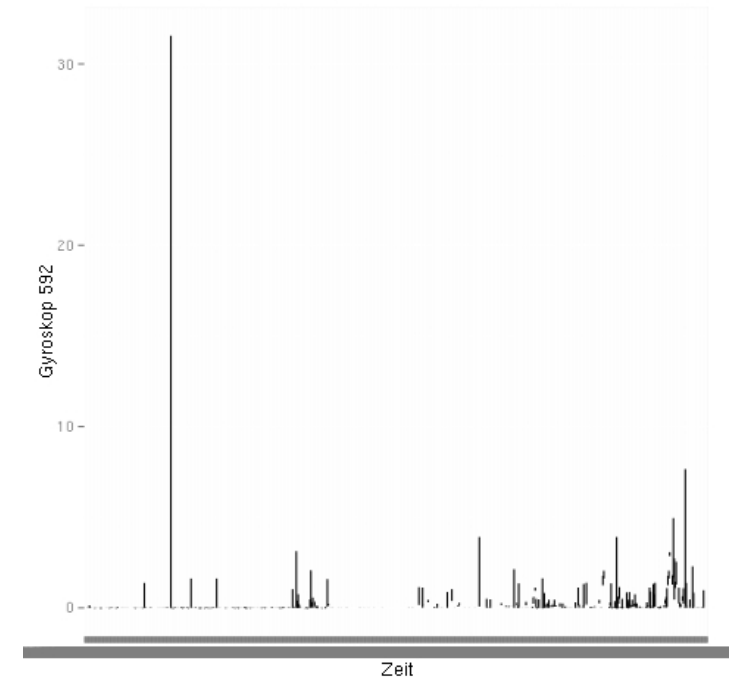


- Not all data is telling straightforward like geolocation. Gyroscope data e.g. is measured in three dimensions.
- This plot shows typical artefacts: the spikes shooting out of the clutter in regular intervals. These are caused by hardware inaccuracies, or also by aliasing effects.
- The artefacts are unique to each device, like a fingerprint, and can identify the source of the data.



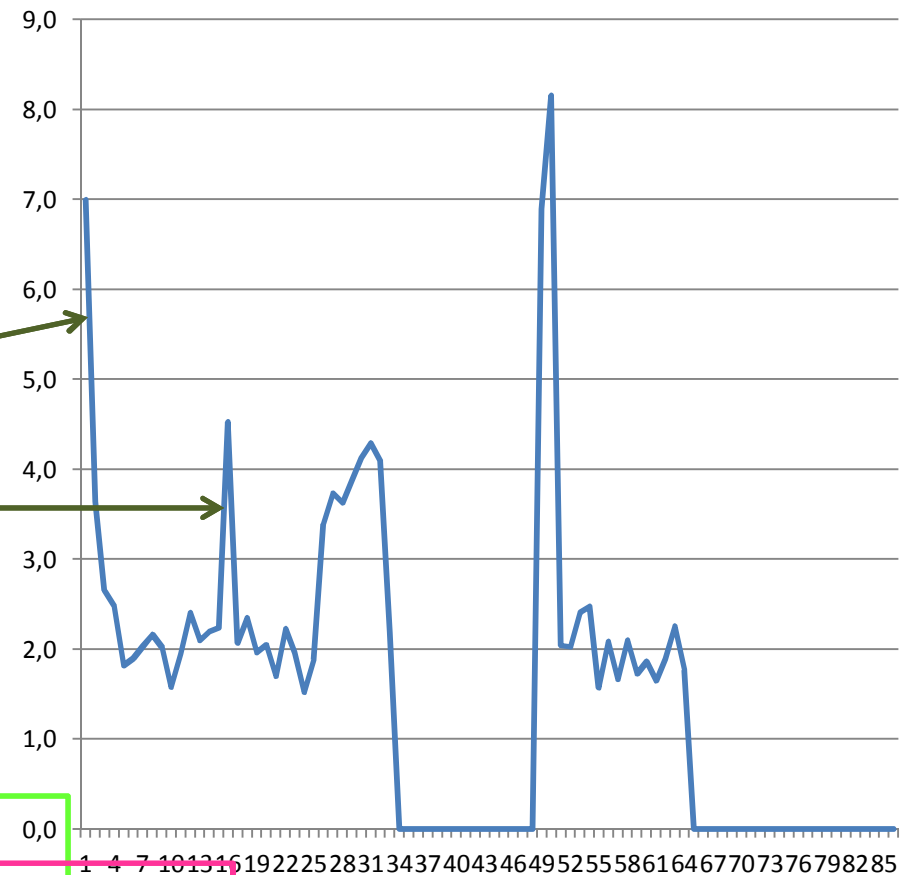
Events

- To see what happens, we have to process the data. How people move around is visible through the gyroscope - you see the turns, changes in directions ect.
- With gyroscopic data in combination with acceleration and speed, also the means of transportation can be revealed: walking has a distinct signature, driving by car shows more changes in directions then sitting on a train, etc.
- However: the data is noisy; artefacts emerge from different brands of the sensors, of glitches in the operating systems, and also can be caused by environmental influences.
- Take e.g. the rhythmic spikes in the picture below: nobody would turn rhythmically and so fast.
- So we have to preprocess the data in the app, to really see, what happens.



What is behavior?

- The normalized gyroscopic data on the right shows the movements of a person going from her desk into the kitchen, fixing a pot of tea, leaving the kitchen and returning to her desk.
- Sampling rate was 10s, timeframe is 15min.
- We notice episodes of different behavior.
- turning sharply
- walking
- turning smoothly
- walking again
- entering the kitchen, preparing the pot
- waiting for the water to boil
- standing up, leaving the kitchen
- sitting down again



Datarella - Joerg Blumtritt

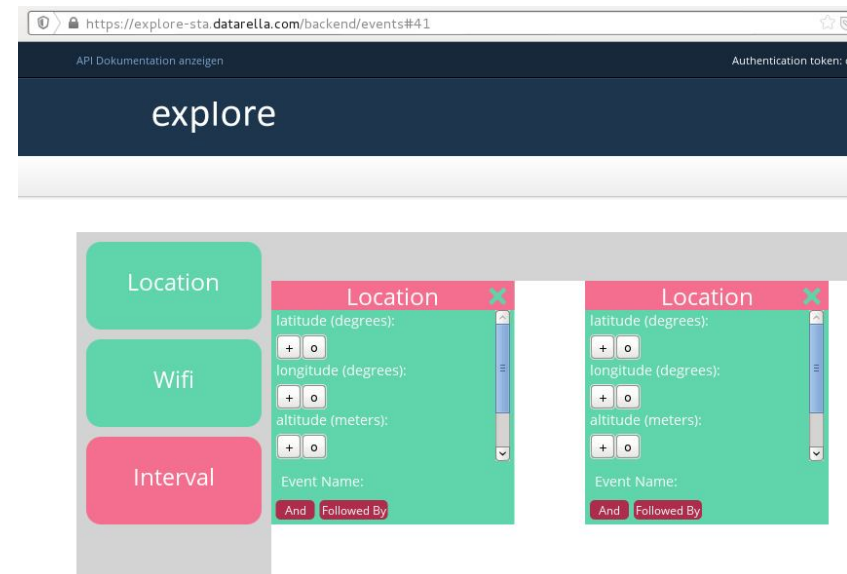


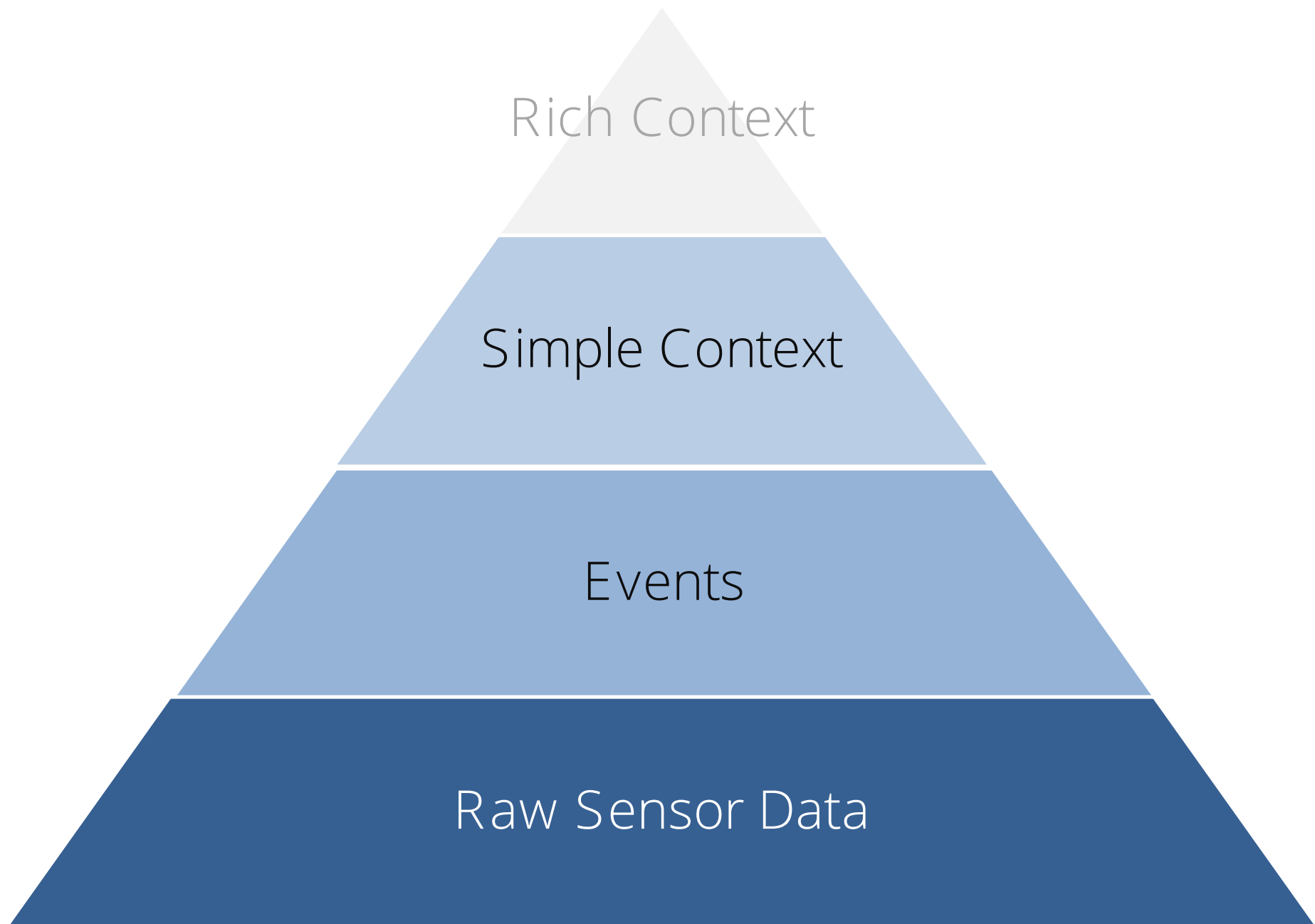
Complex Event Processing

- Simple events, like changing direction, entering an area of specific geo-coordinates, or having moved for a specific time span can be combined to complex events.
- EPL (event processing language) offeres a way to listen to the data stream and detect the occurance of events.
- EPL looks like SQL, but instead of tables, the search goes into the data stream.
- For our app, we define events, boolean-combine these events in a GUI and parse the definition in the app via JSON doc.
- The event processing itself takes place in the app - no network connection is needed.



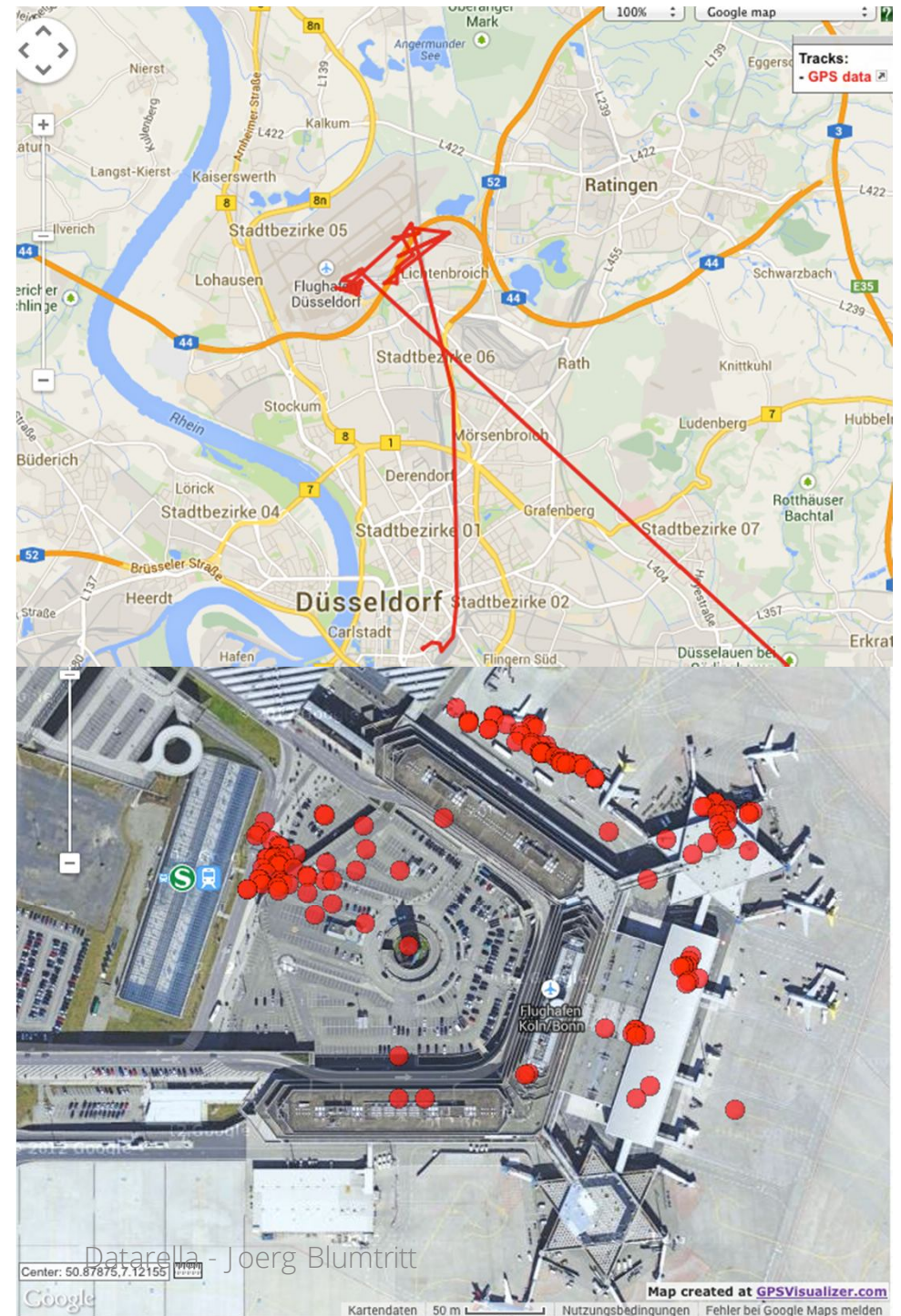
```
SELECT ID AS sensorId  
FROM ExampleStream  
RETAIN 60 SECONDS  
WHERE Observation= " Outlet"
```

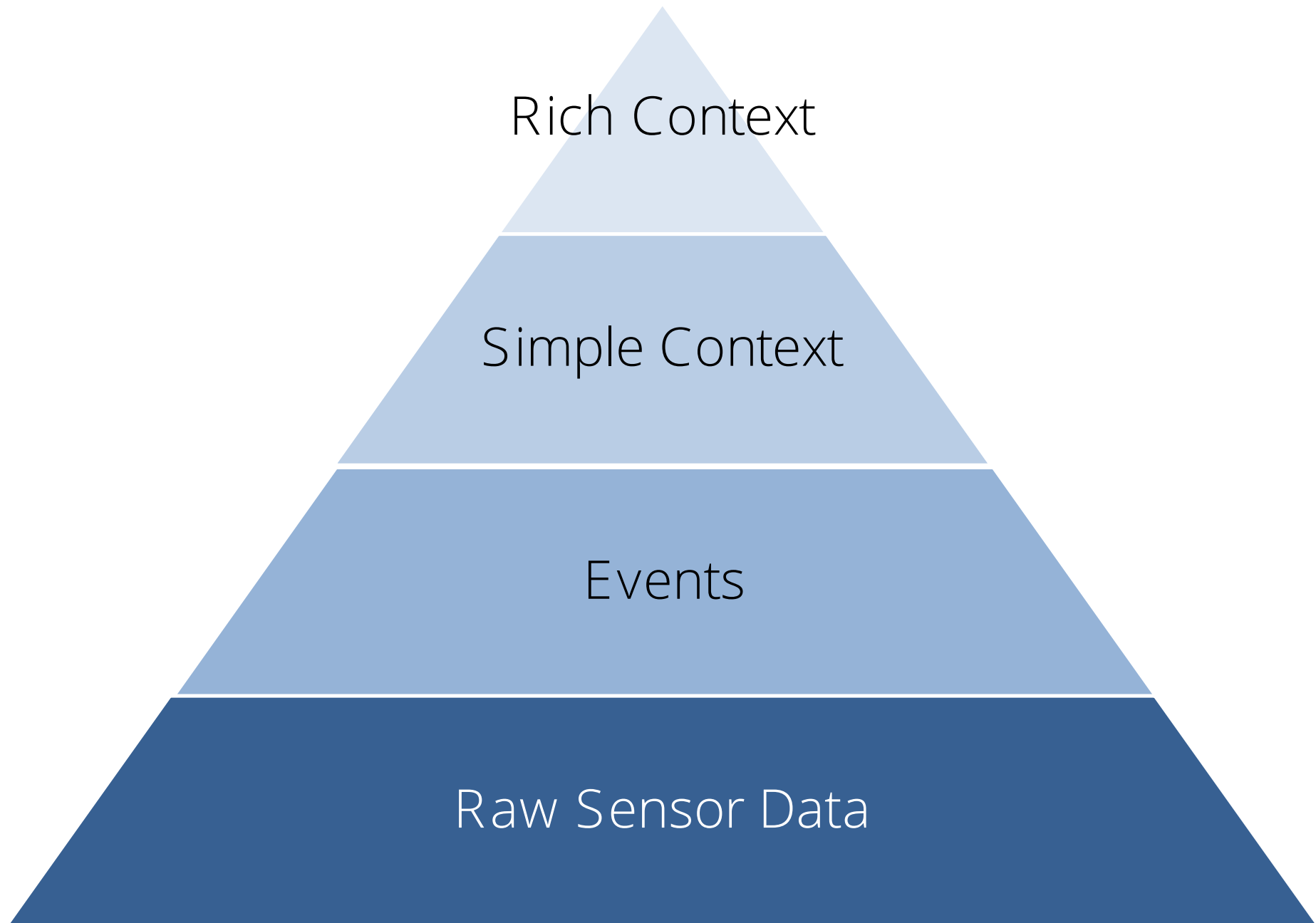




Travel

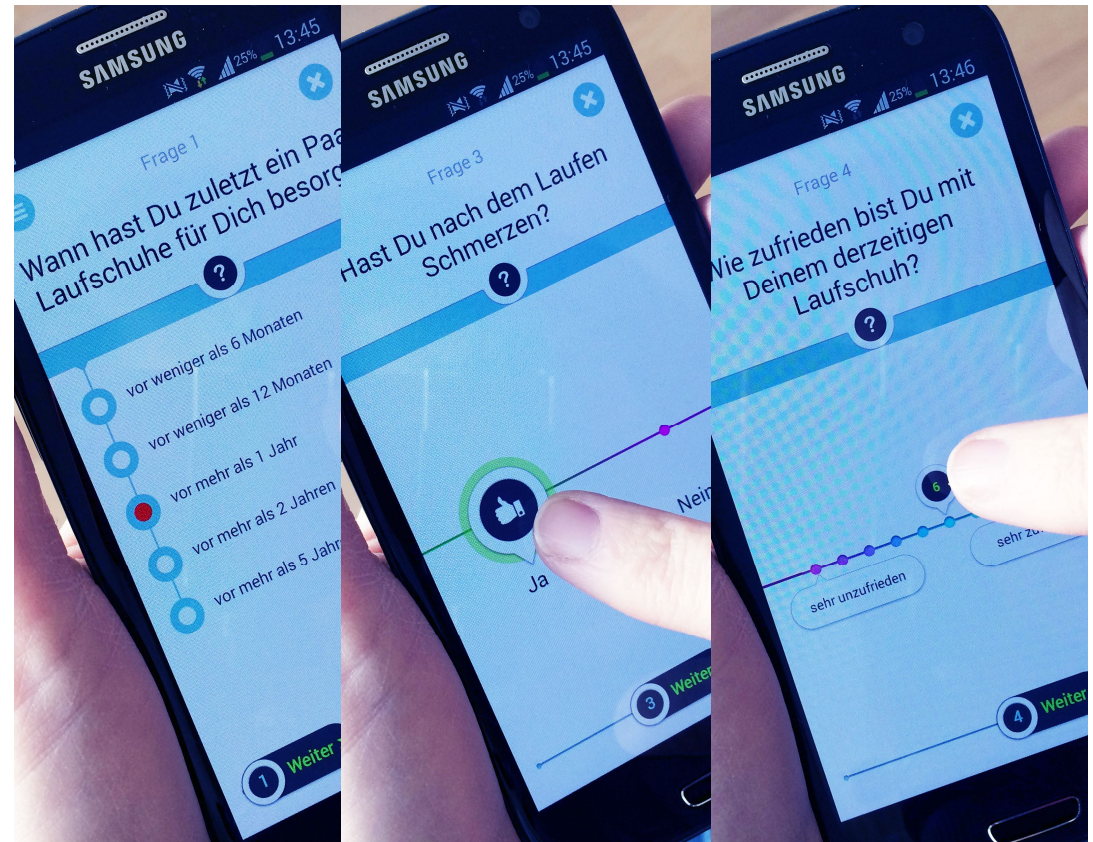
- Studying the means of transportation, the paths people choose for their commute or travel, is a straightforward application of our data.
- We work e.g. for airports to optimize the shops they would offer to passengers. Since many passengers come from other cultures, it is not an easy task for an airport (or in general for a shopping mall) to learn the preferences of potential clients - not consistent shopping data or market research is available.
- So, e.g. we incentivize passengers from China to let us accompany their stay in Europe with our app 'explore'. So we can understand, what they wanted to buy, if they succeeded and if they would have missed anything, that an airport could have offered to them.





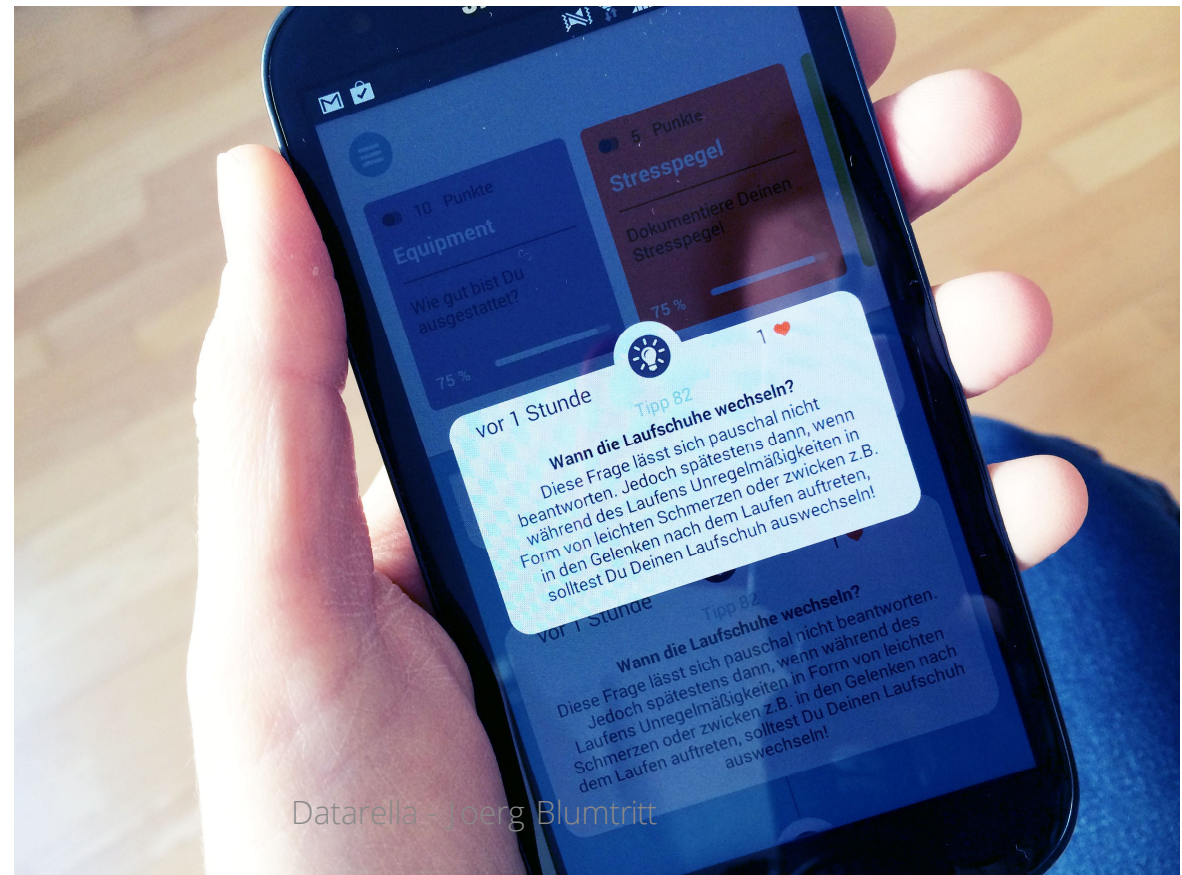
Usability

- To avoid the shortcomings of the other social research apps, we focused on the user interface, to make it as "mobile" as possible.



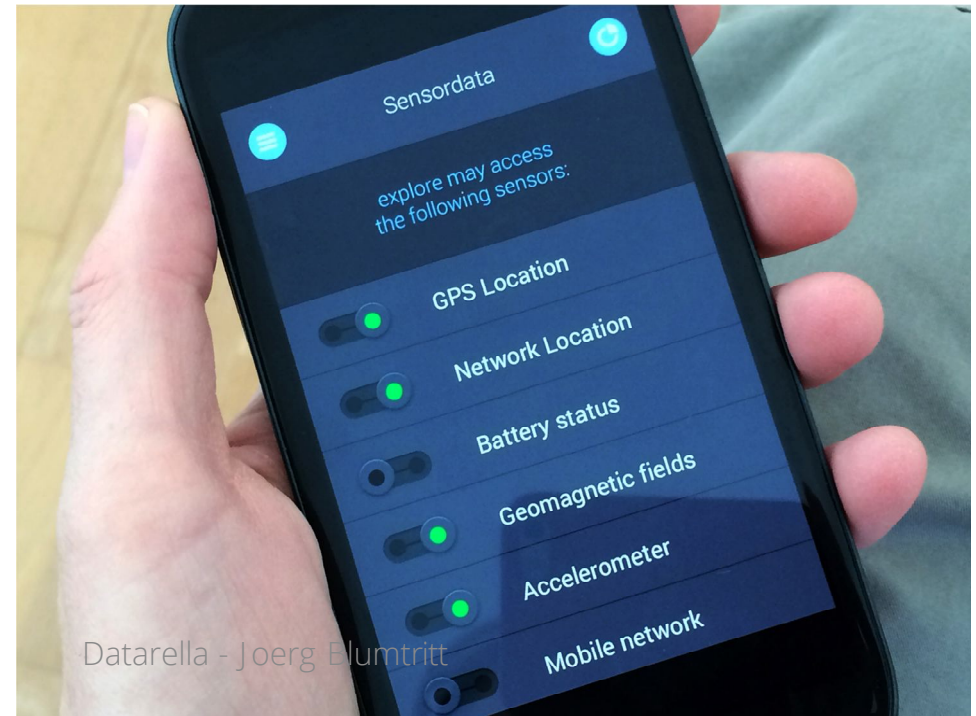
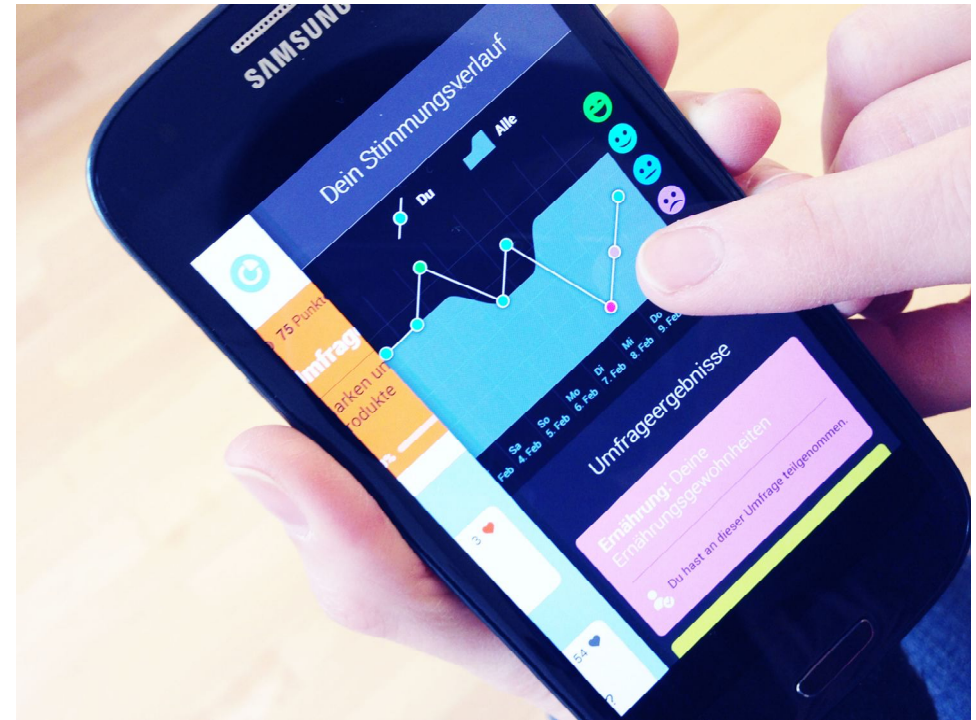
Interactions

- explore can ask questions or offer suggestions, triggered by data.

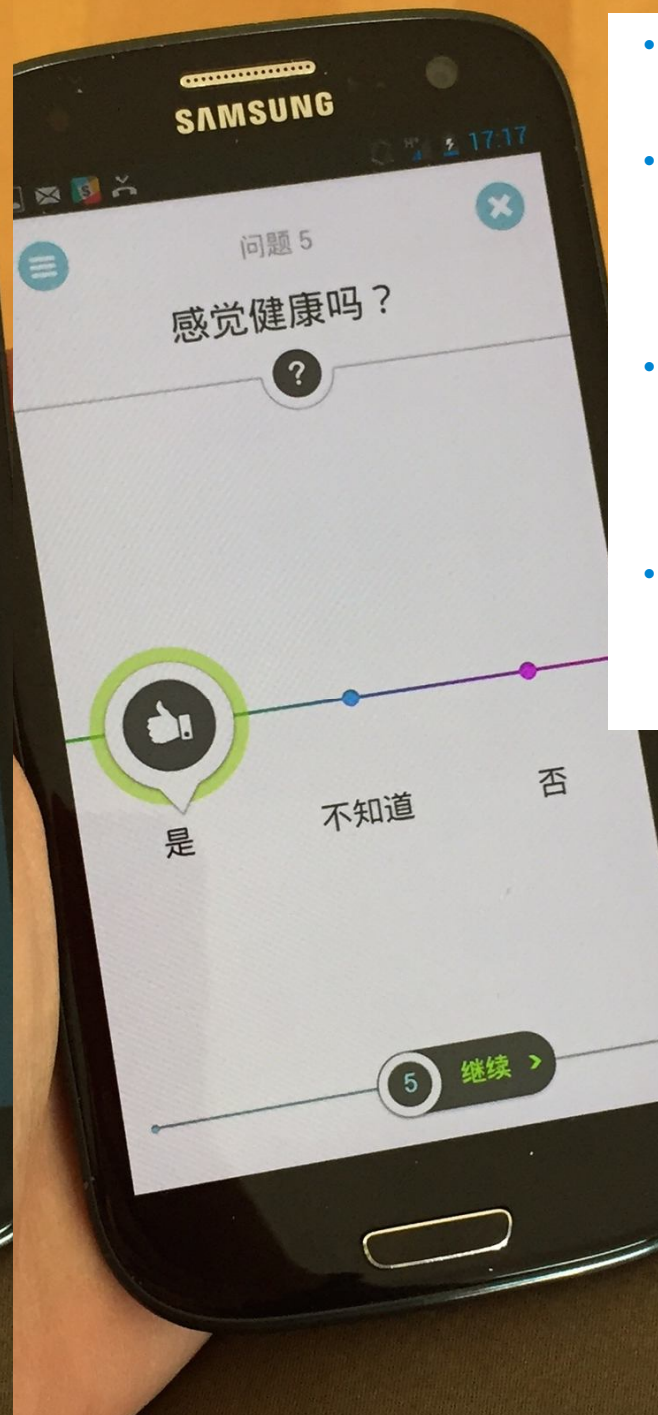
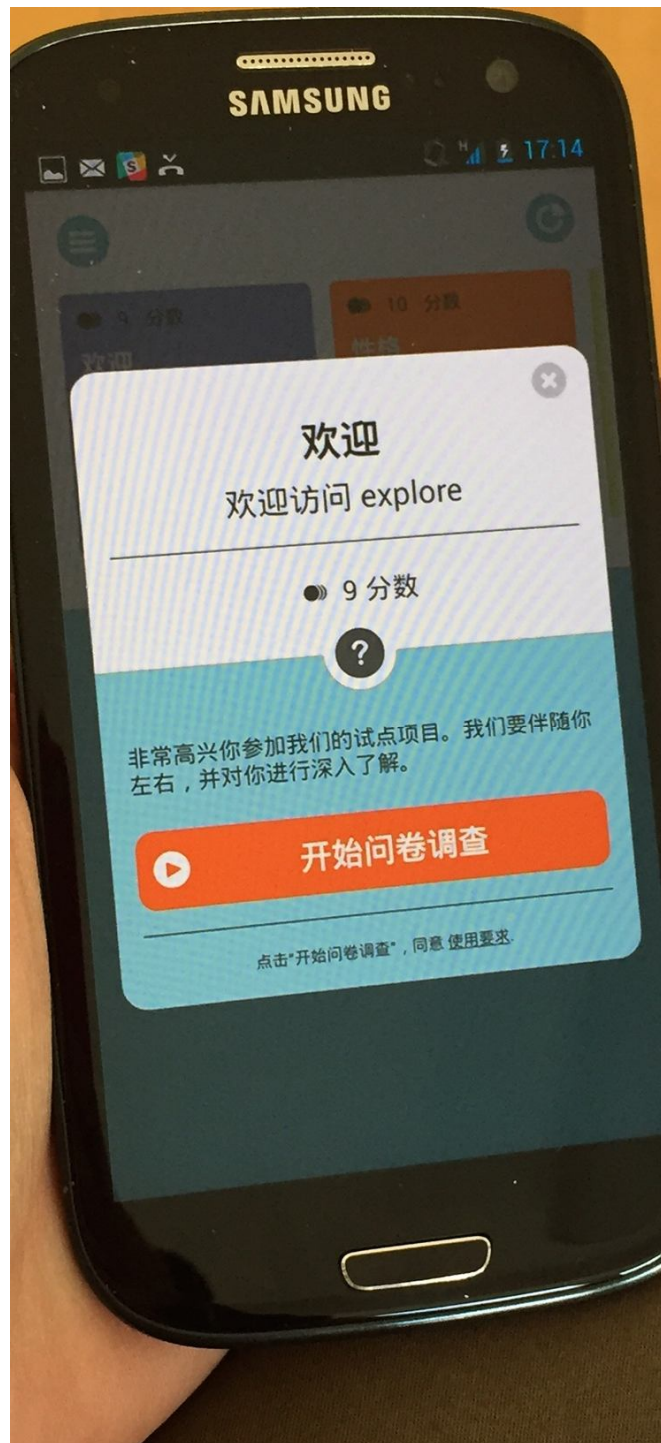


Quantified Self

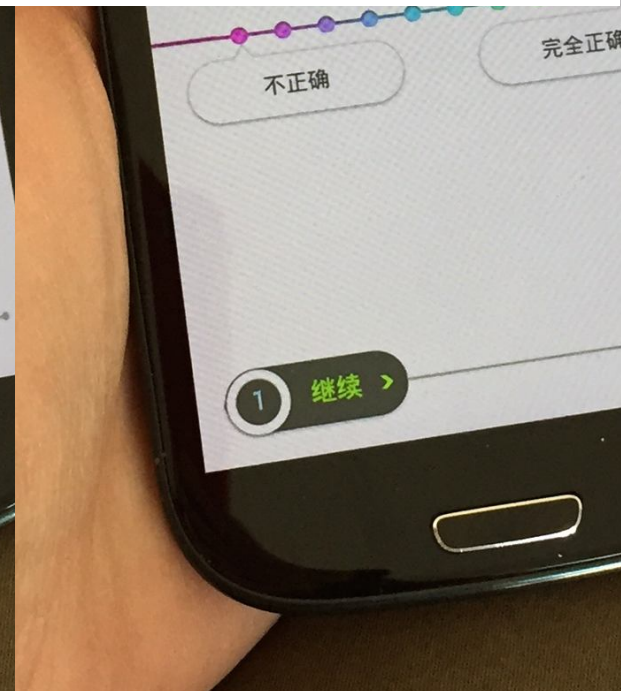
- explore offers clear and simple analytics of the data collected. People can also get their raw-data for their own purposes.
- We want people to be aware what we (and other apps) do on the phone. So we do not only tell in advance, we also show what sensors are activated and give the opportunity to opt-out per sensor.
- Since we reflect the results of our tracking as well as questionnaires and interactions in form of diagrams and summaries, we hope, people will realize what we are doing and can act self-determined.
- Of course we respect take-down notices: if people ask for their data to be deleted, we follow their request (which btw is also required by German data protection laws); this is also a reason for us not to use common cloud storage and cloud computing platforms, since we would not have control over the back-up.

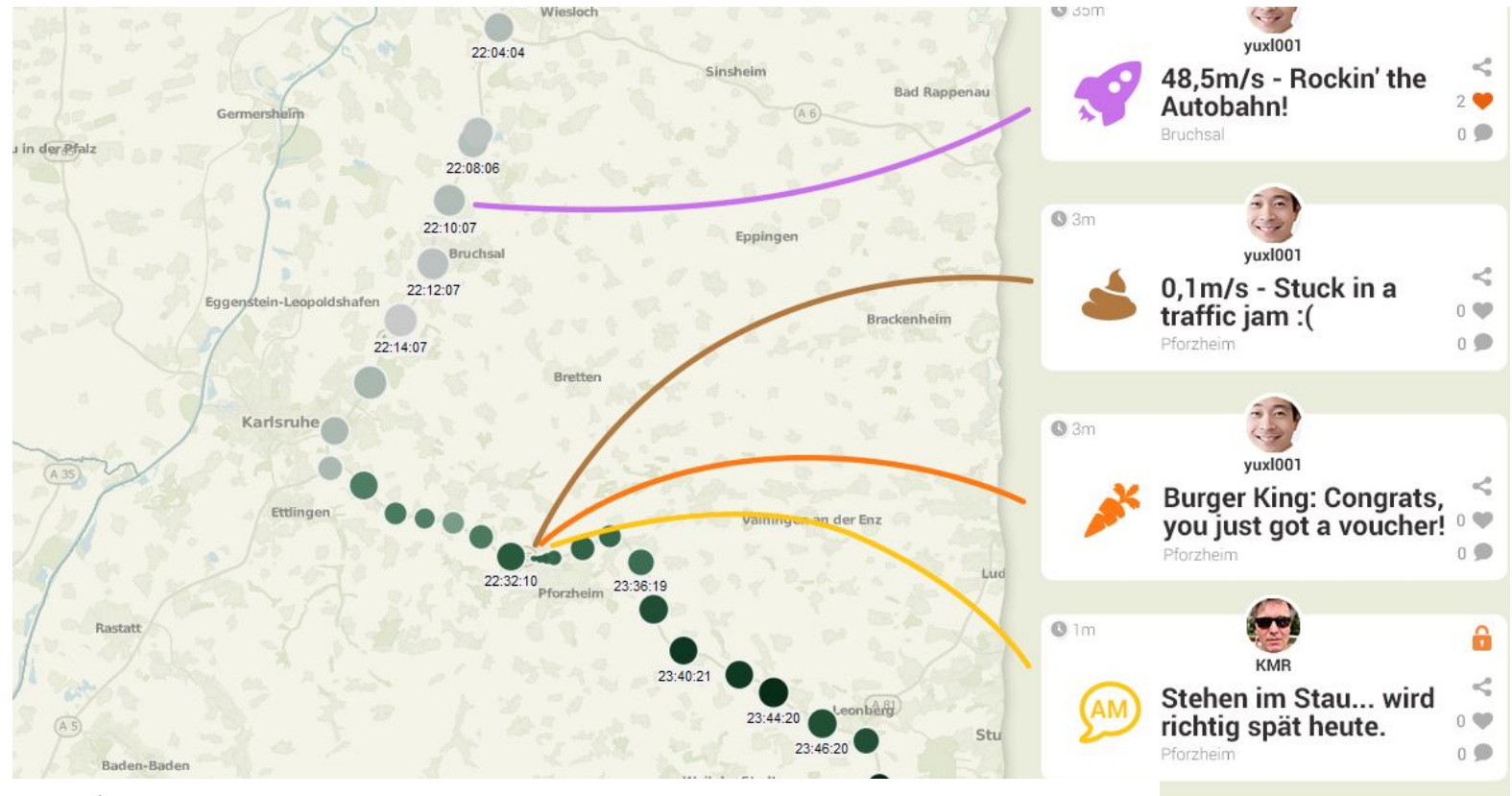


Datarella - Joerg Blumtritt



- Use case: Tracking Chinese Passengers.
- An international airport wants to learn what Chinese passengers buy, what they consider a pleasant shopping experience, and what they expect.
- We recruited a panel of Chinese passengers before they left from China to Europe, and accompanied them with our app.
- We learned where they went and could ask them about their experiences.

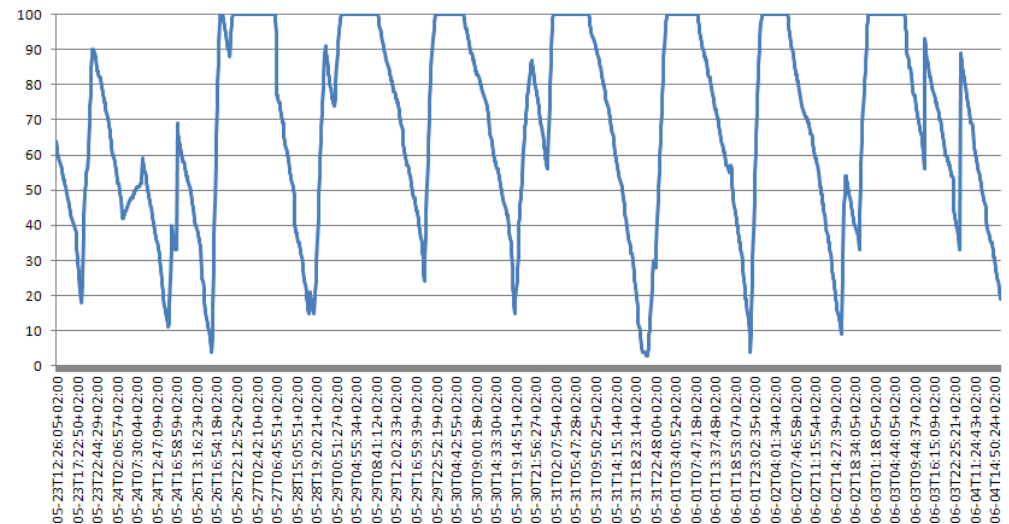




- Use case: Driver Timeline
- We built an app that tracks the driving and computes more abstract events from its data, like „stuck in a traffic jam“.
- The events are displayed in form of a timeline, and can be shared to others.

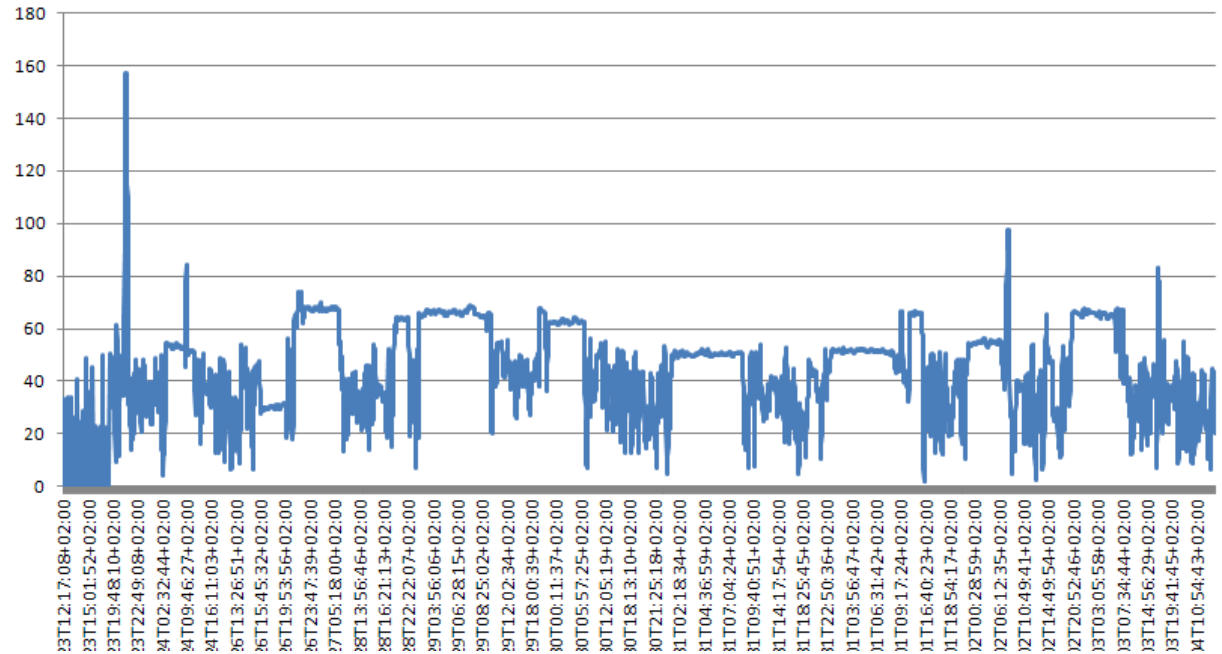
Battery

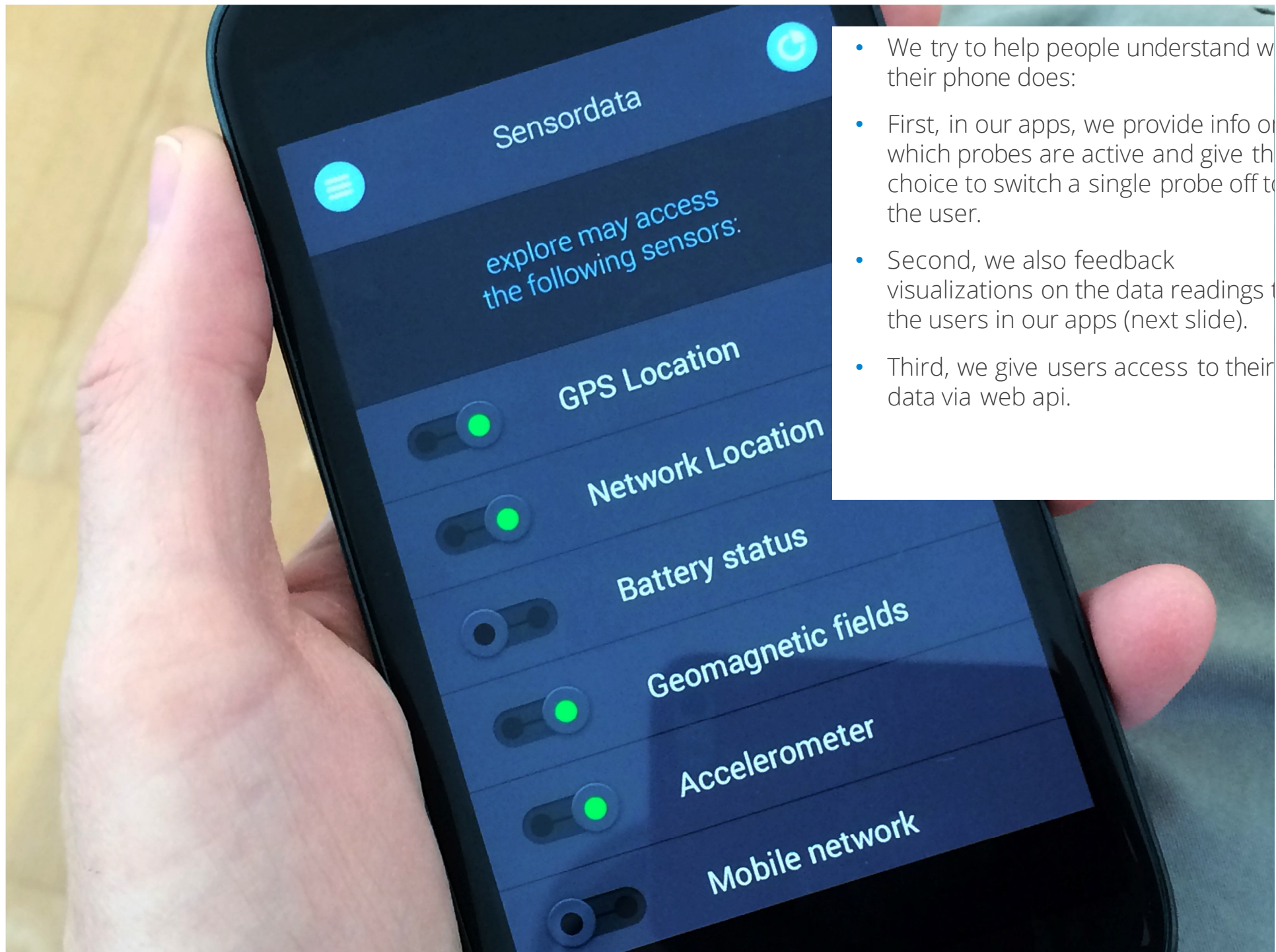
- Battery data is both interesting in itself, and also important to maintain the app usable.
- Battery consumptions is telling a lot about the environment of the phone: temperature, moisture, even air pressure can be derived using the change in charge.



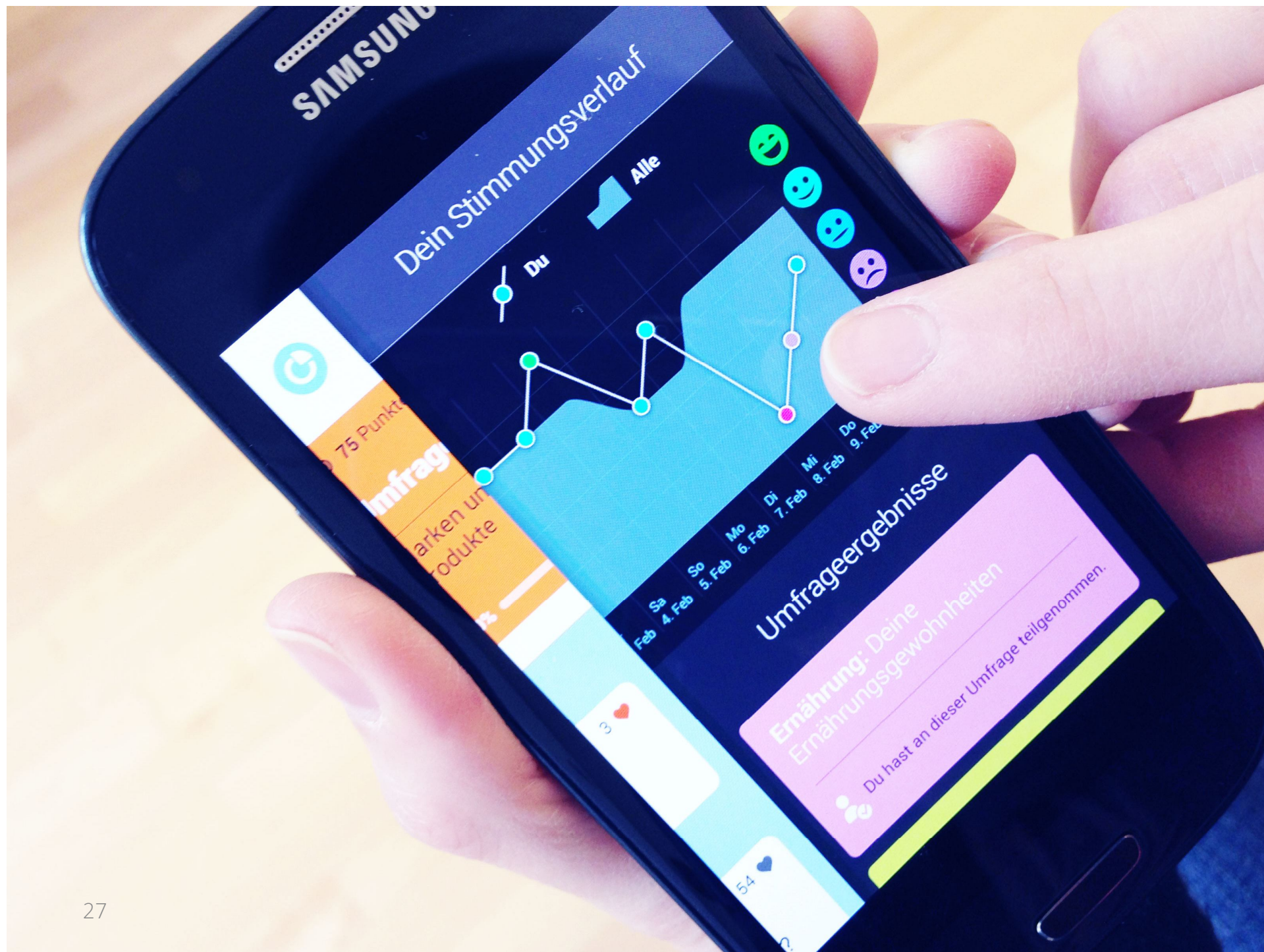
Identify whereabouts by location-specific magnetic field

- Every place has a distinct signature of the magnetic field (in strenght like shown on my own tracking data on the right as well as in bearing).
- So even if someone decided to not-track geolocation, we might still get sufficient information on their whereabouts via other measurements.
- That this is not hypothetical can be seen on the diagram: the field's signature of my home is different from other places, I stayed during that week.





- We try to help people understand what their phone does:
- First, in our apps, we provide info on which probes are active and give the choice to switch a single probe off to the user.
- Second, we also feedback visualizations on the data readings to the users in our apps (next slide).
- Third, we give users access to their data via web api.



Probes

Authentication

Authentication is done via OAuth. Use the e-mail and password you are using with the Explore App to authenticate.

```
curl -F grant_type=password \  
-F username=yourApp@email.de \  
-F password=yourPassword \  
-F scope=user \  
-X POST https://explore.datarella.com/api/oauth/token.json
```

Output:

```
{"access_token":"7df51cc6dcfbeaf56bffa18e4dd4d6acd85e83f377f5bee4d3ebac43e3872db33","token_type":"bearer","expires_in":7200,"scope":"user"}
```

This access_token is used in all following requests. You need to pass this token in the HTTP-Header as Authorization: Bearer :token as described in self-issued.info/docs/draft-ietf-oauth-v2-bearer.html#authz-header Wrong login credentials will produce the HTTP status code 401 Unauthorized

For example

```
curl -i -H "Accept: application/json" -H "Authorization: Bearer $access_token" -X GET https://explore.datarella.com/api/data/v1/...
```

GET /api/data/v1/probes

>>>

Get a list of probes

GET /api/data/v1/probes/:probe_type

>>>

Get a list of probe samples

GET /api/data/v1/probes/:probe_type/:id

>>>

Get a specific probe sample

GET /api/data/v1/probes/aggregated/:probe_type

>>>

Get a list of aggregated probes

Spooky Wifi

Self-Tracking vs. Others-Tracking

- You can't avoid tracking others involuntarily, too. This is a problem. People might be aware, what they themselves are doing. But others might be tracked along without giving their consent.
- Wifi is a good example of "others-tracking": all wifi signals within reach are tracked by the phone. It tells a lot about other people; not only about the devices they use.

bssid	ssid	capabilities	frequency	level	created_at
1c:c6:3c:31:e0:f9	Laperla	[WPA2-PSK-CCMP][WPS][ESS]	2462	-95	2014-05-23T12:00:00
5c:35:3b:7b:f8:29	KDG-BF824	[WPA-PSK-TKIP+CCMP][WPA2-PSK-TKIP+CCMP][WPS][ESS]	2412	-88	2014-05-23T12:00:00
44:32:c8:35:3d:ff	Tech_D0053811	[WPA-PSK-TKIP+CCMP][WPA2-PSK-TKIP+CCMP][ESS]	2437	-94	2014-05-23T12:00:00
24:65:11:e7:5a:0f	FRITZ!Box 7312	[WPA-PSK-TKIP][WPA2-PSK-CCMP][WPS][ESS]	2412	-89	2014-05-23T12:00:00
88:25:2c:64:e0:1e	WLAN-64E037	[WPA-PSK-TKIP+CCMP][WPA2-PSK-TKIP+CCMP][WPS][ESS]	2437	-92	2014-05-23T12:00:00
84:1b:5e:01:73:22	Jagertee Ischgl	[WPA-PSK-TKIP+CCMP][WPA2-PSK-TKIP+CCMP][WPS][ESS]	2437	-95	2014-05-23T12:00:00
00:1a:2a:c5:2a:20	WLAN-Koebisch	[WPA-PSK-TKIP+CCMP][WPA2-PSK-TKIP+CCMP][WPS][ESS]	2427	-78	2014-05-23T12:00:00
50:7e:5d:16:6a:85	root	[WPA-PSK-TKIP+CCMP][WPA2-PSK-TKIP+CCMP-preauth][WPS][ESS]	2467	-60	2014-05-23T12:00:00
44:94:fc:58:bd:50	NETGEAR57	[WPA2-PSK-CCMP][WPS][ESS]	2437	-83	2014-05-23T12:00:00
34:6b:d3:43:1f:67	WLAN-1f6765	[WPA2-PSK-CCMP][WPS][ESS]	2442	-90	2014-05-23T12:00:00
24:65:11:9c:fa:46	FRITZ!Box WLAN 3270	[WPA-PSK-TKIP][WPA2-PSK-CCMP][WPS][ESS]	2412	-81	2014-05-23T12:00:00
1c:c6:3c:48:e3:cc	ALICE-WLAN21	[WPA2-PSK-CCMP][WPS][ESS]	2467	-85	2014-05-23T12:00:00
34:31:c4:17:c6:90	WebbNet	[WPA2-PSK-CCMP][WPS][ESS]	2412	-85	2014-05-23T12:00:00
bc:05:43:d3:2d:37	philotas	[WPA-PSK-TKIP][WPA2-PSK-CCMP][WPS][ESS]	2442	-82	2014-05-23T12:00:00
00:1e:f9:e8:88:f0	Thom_D011882	[WPA-PSK-TKIP+CCMP][WPA2-PSK-TKIP+CCMP][ESS]	2462	-85	2014-05-23T12:00:00
04:35:3b:7b:f8:2b	KD WLAN Hotspot+	[ESS]	2412	-91	2014-05-23T12:00:00
00:1d:19:e3:29:88	Arcor-E32912	[WPA-PSK-TKIP+CCMP][WPA2-PSK-TKIP+CCMP-preauth][ESS]	2452	-88	2014-05-23T12:00:00
44:32:c8:9c:28:21	Tech_D0048747	[WPA-PSK-TKIP+CCMP][WPA2-PSK-TKIP+CCMP][ESS]	2412	-78	2014-05-23T12:00:00

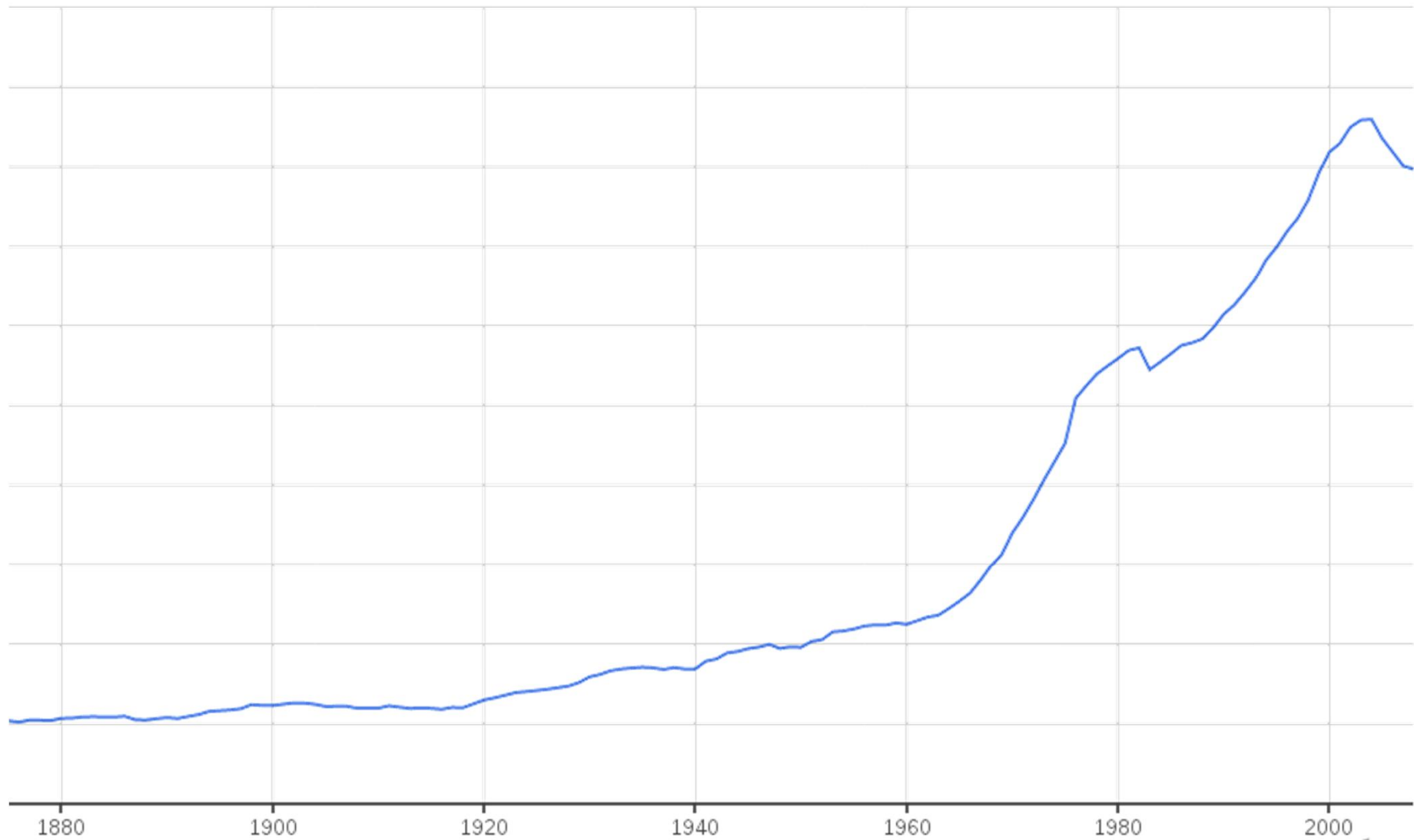
royal	[WPA-PSK-TKIP][WPA2-PSK-CCMP]
HP-Print-c2-LaserJet 100	[ESS][P2P]
royal	[WPA-PSK-TKIP][WPA2-PSK-CCMP]
HP-Print-c2-LaserJet 100	[ESS][P2P]
royal	[WPA-PSK-TKIP][WPA2-PSK-CCMP]
HP-Print-c2-LaserJet 100	[ESS][P2P]
royal	[WPA-PSK-TKIP][WPA2-PSK-CCMP]
HP-Print-c2-LaserJet 100	[ESS]
royal	[WPA-PSK-TKIP][WPA2-PSK-CCMP]
HP-Print-c2-LaserJet 100	[ESS][P2P]

Postprivacy, and communalization of private life

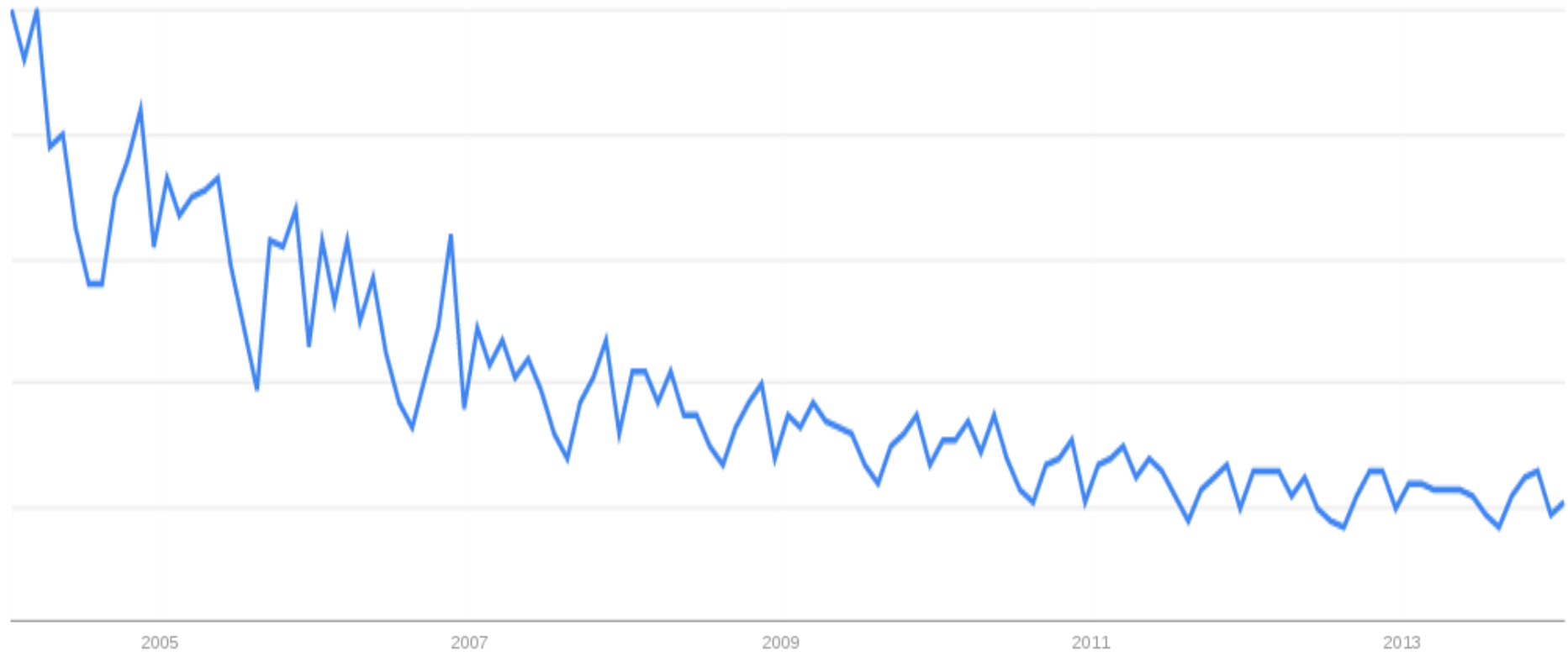
- In Neal Stephenson's "Snow Crash", we read about the 'Central Intelligence Corporation' - a commercialized version of today's NSA. Mobile health, computational social science, and mass measurement of environmental influences are obvious and benign applications of QS for the public good. With quantifying and making public, what European data protection law defines as "the most intimate personal data", however do we transform the current "knowledge-database" character of the Net along with its communication-networks to something new, something that might become similar to Stephenson's vision?
- Could this even lead to Teilhard's (resp. McLuhan's) angelization of humans, not only connected via social media but bodily knit into the data? Would we rather end up in a rally bucolic global village with moral control by the panoptic community (and an inherent abelism that comes with a village life)? In both aspects, representative aggregates like society as well as the concept of the individual might be rendered obsolete.



"Privacy"



"Data Protection"



Becoming cyborgs?

- The bodily extension into the data square - this is what cyborgism is really about.
- People like Neal Harbison or Enno Park are pushing the discussion in that direction: How do we maintain possession of our bodies? What ethic framework has there to be set-up? How do we avoid technological extensions becoming "black boxes" that control us, rather than we do them?
- So it is worthwhile to follow the proceedings of the Cyborg e.V that Enno founded.



079: Hörspiele, Schweine und Maschinenmenschen



Some links

- <http://datarella.com/blog>
- <http://beautifuldata.com>
- <http://twitter.com/jbenno/bigdata>

my blogs.

a data science related twitter list.



Jörg Blumtritt
@jbenno

Datarella GmbH
Oskar-von-Miller-Ring 36
80333 München

089/44 23 69 99
info@datarella.com



DATARELLA

@jbenno