

Interactive Feedback of Data Quality in Clinical Research

A Case Study from an Infectious Diseases Cohort

Stephan Glöckner

email: stephan.gloeckner@helmholtz-hzi.de

Osnabrück, September 3rd, 2018

Outline

1. Feedback of Data Quality
2. TxCohort
3. Developed Dashboard framework
4. Limitations/Conclusions

1. Feedback of Data Quality

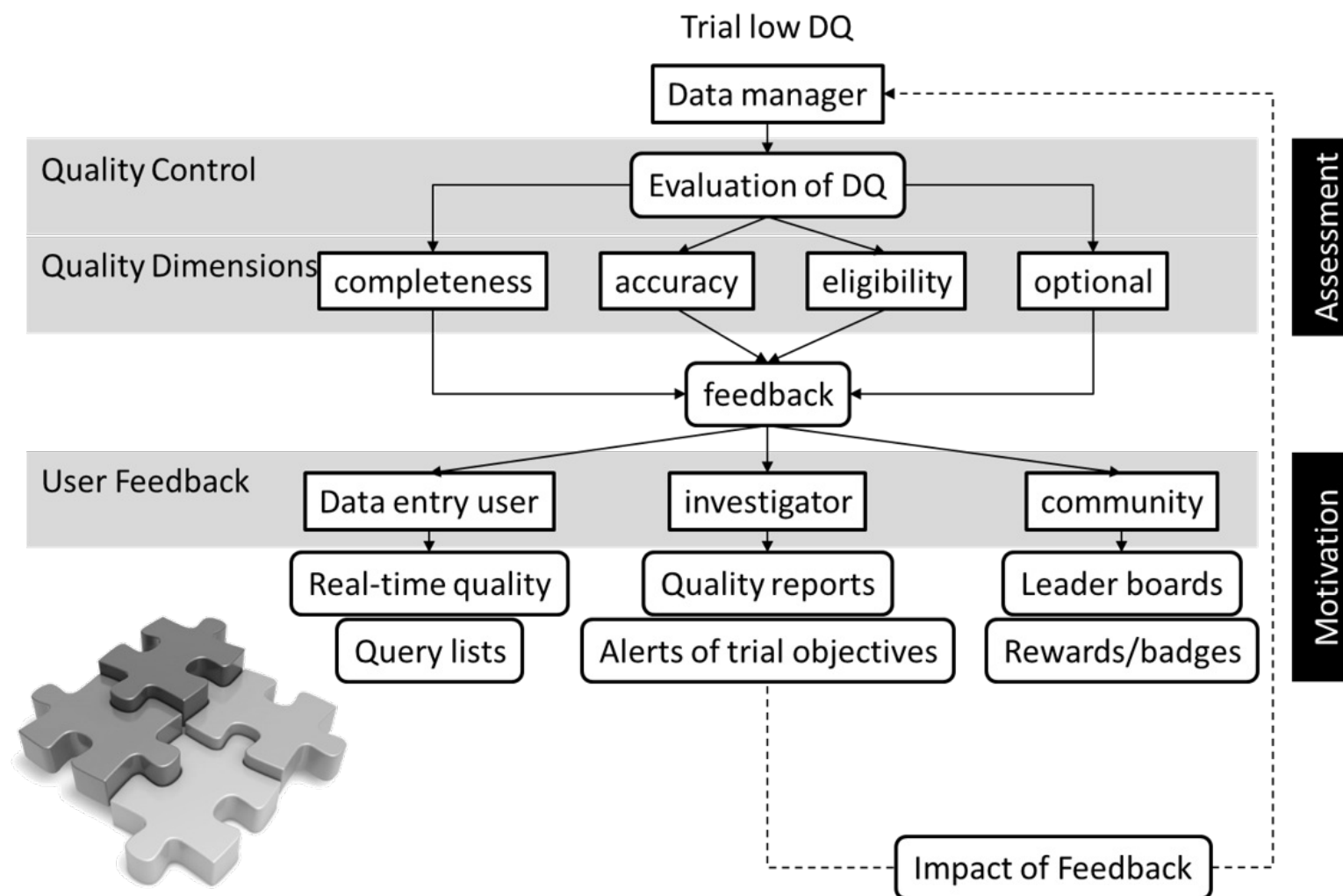
Evaluation of Data Quality

	A: COMPLETE	B: CORRECT	C: CURRENT
1: PATIENTS	1A There are sufficient data points for each patient.	1B The distribution of values is plausible across patients.	1C All data were recorded during the timeframe of interest.
2: VARIABLES	2A There are sufficient data points for each variable.	2B There is concordance between variables.	2C Variables were recorded in the desired order.
3: TIME	3A There are sufficient data points for each time.	3B The progression of data over time is plausible.	3C Data were recorded with the desired regularity over time.

Note: Data quality constructs are at the top, data dimensions along the side, and cells contain corresponding operationalized constructs.

1. Feedback of Data Quality

Data Quality Feedback



1. Feedback of Data Quality

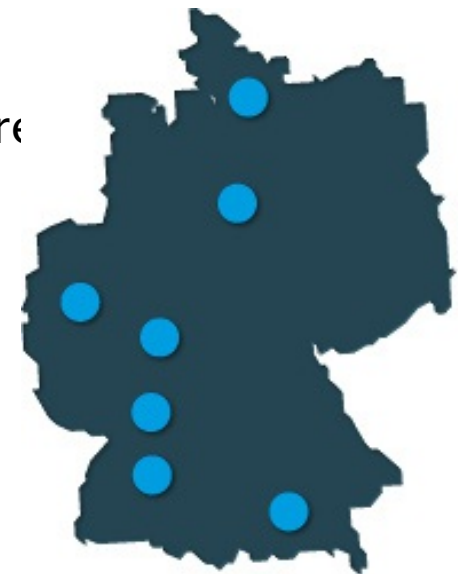
Interactivity needed?

- in some way, complex/high dimensional data has to be reported
- **busy** clinicians/investigator need all information on the **first look**
- on demand: **filtering, selecting, zooming**
- dynamic, simple, accessible --> **web-based & real-time applications**

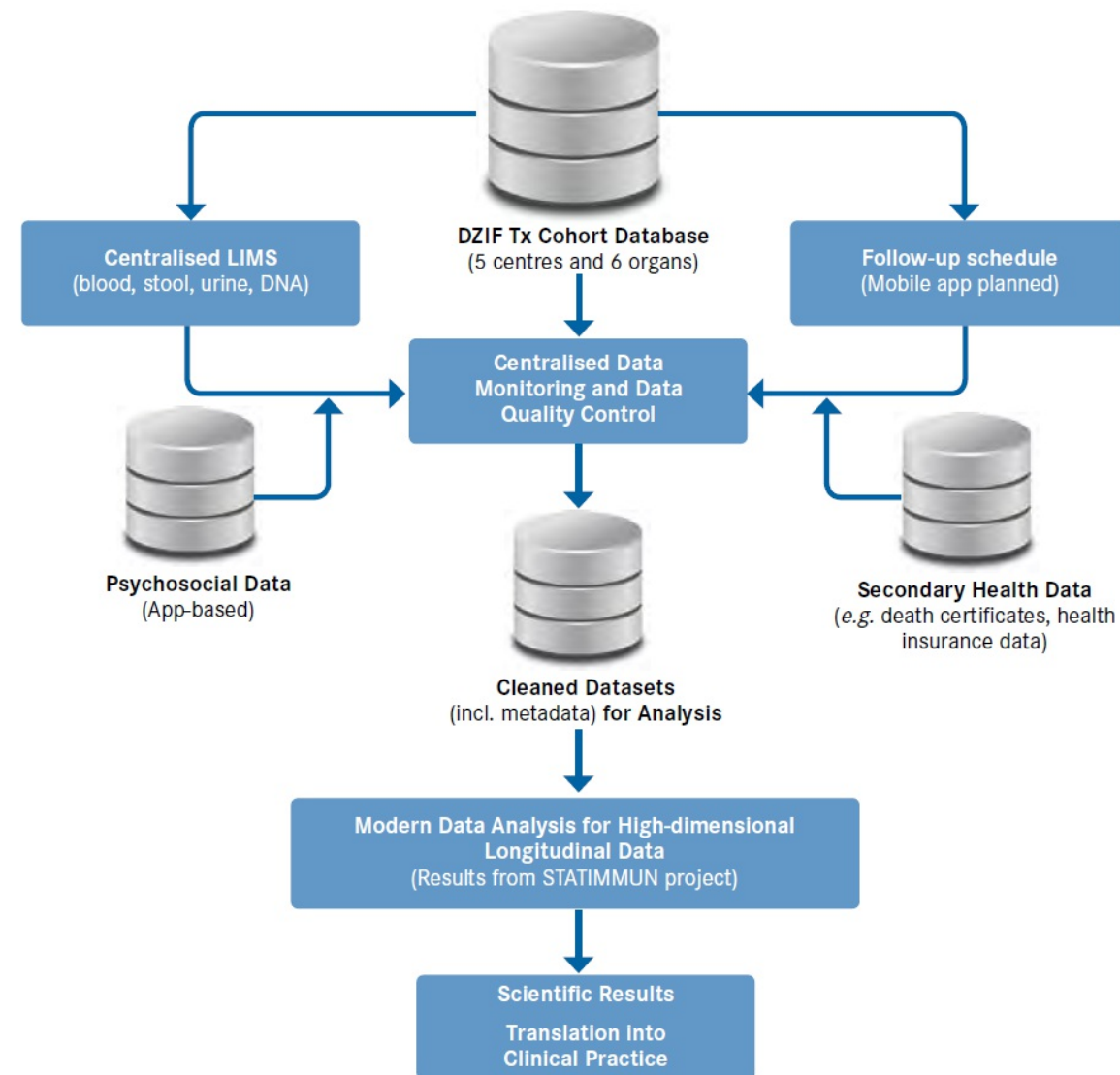
2. TxCohort

Transplantation Cohort – an Infrastructure of the German Center for Infection Control (DZIF)

- multicenter study with ~600 participants (all ages)
- participants after transplation of heart, lung, pancreas, kidneys, liver or stem cells
- **primary aim:** organ transplantation and its influence on infection susceptibility and organ function
- 14 heterogene patient records (demography, postoperative procedure, bacterial infections, etc.)
- monthly exports with in separate csv files (5 centers x 6 organs x 14 records)



2. TxCohort – Data Processes



3. TxDashboard – Needs Assessment

- Study Nurses, Clinicians, (Local-) Investigators
- content:
 - Study Population / Recruitment Rate
 - Biosamples
 - Follow-up Performance
 - Data Quality
- single page, quick to understand
- common scripting language researcher and clinician (can) understand:
 - **language:** R
 - **documentation:** rmarkdown
 - **graphs:** ggplot2, googleVis, leaflet, plotly
 - **tables:** knitr, kableExtra, DT, formattable
 - **color-scales:** viridis, colorbrewer: RdYlBu

Auf einem Blick

9723

Blutproben



4277

Andere Proben



1998

Patienten mit Bioproben



Blutproben

TOTAL

Niere

Leber

Lunge

Herz

Pankreas

Stammzellen

Copy

CSV

Search:

	sample_type	Center_1	Center_2	Center_3	Center_4	Center_5	TOTAL
1	EDTA-Buffy Coat	248	237	227	270	209	1191
2	EDTA-Plasma	251	228	234	238	235	1186
3	EDTA-Primärröhrchen	246	231	238	234	230	1179
4	keine Angabe	234	235	240	225	270	1204

Showing 1 to 9 of 9 entries

Previous

1

Next

Andre Proben

TOTAL

Niere

Leber

Lunge

Herz

Pankreas

Stammzellen

Copy

CSV

Search:

	sample_type	Center_1	Center_2	Center_3	Center_4	Center_5	TOTAL
1	Bukkaler Abstrich	156	170	178	171	165	840
2	Rachenspülwasser	152	169	173	180	178	852

Showing 1 to 6 of 6 entries

Previous

1

Next

Übersicht Niere FU

Übersicht Leber FU

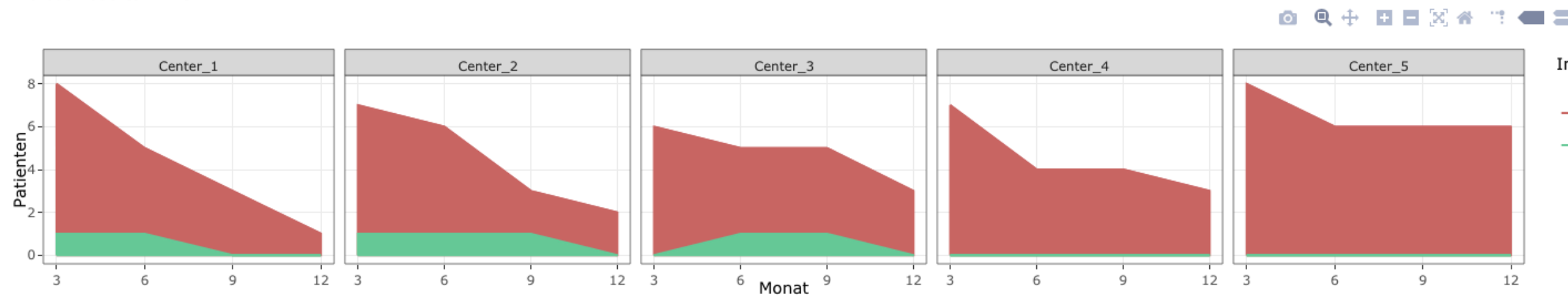
Übersicht Lunge FU

Übersicht Pankreas FU

Übersicht Stammzellen FU

Übersicht Herz FU

Anzahl der Patienten im FU



Übersicht Niere FU

Übersicht Leber FU

Übersicht Lunge FU

Übersicht Pankreas FU

Übersicht Stammzellen FU

Übersicht Herz FU

Show 10 entries

Search:

Center	Month 3			Month 6			Month 9			Month 12		
	ACTUAL	TARGET	in %	ACTUAL	TARGET	in %	ACTUAL	TARGET	in %	ACTUAL	TARGET	in %
Center_1	1	8	12.5	1	5	20	0	3	0	0	1	0
Center_2	1	7	14.29	1	6	16.67	1	3	33.33	0	2	0
Center_3	0	6	0	1	5	20	1	5	20	0	3	0
Center_4	0	7	0	0	4	0	0	4	0	0	3	0

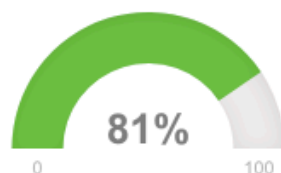
Showing 1 to 5 of 5 entries

Previous 1 Next

Daten - Probleme

Übersicht Patienten ohne Tx und FU Datum

Datenqualitaet Generelle Informationen

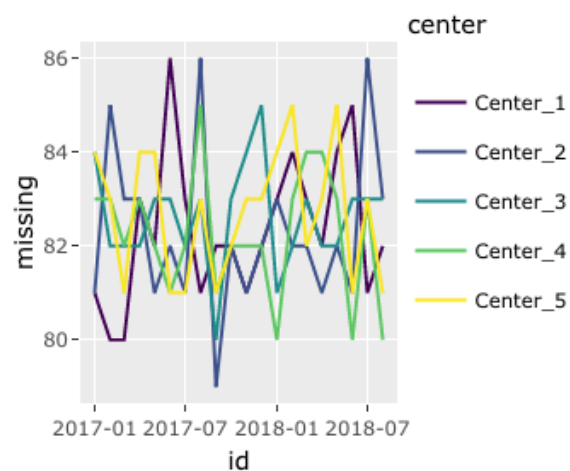


Ranking

x

Center_1	81
Center_2	83
Center_3	80
Center_4	82
Center_5	81

Datenqualitaet ueber Zeit



Variable pro Center

Variable	Center_1	Center_2	Center_3	Center_4	Center_5
Item_1	78	86	82	80	81
Item_2	85	84	78	85	81
Item_3	82	85	84	89	85
Item_4	78	76	77	88	79
Item_5	86	84	78	78	83
Item_6	82	84	78	75	81
Item_7	75	83	80	76	78
Item_8	87	86	85	85	84
Item_9	80	80	76	83	80
TOTAL	81	83	80	82	81
TOTAL	81	83	80	82	81

- Generell
- Post-op
- Bakterielle Infektionen

4. 4. Limitations/Conclusions?

Limitations

- Extension of data quality assessment (more dimensions & indicators in regards to user needs)
- implementation of gamified features (leaderboards, awards, badges)
- check lists (case-based) for study nurses

Conclusions

- busy investigators need fast, individualized feedback
- dashboards provide a interactive solution, but `angular.js` dashboards need to be considered
- data transformation still the most cumbersome process
- DQ needs to be extended and visualization needs improvement (UXD, UXR)

Thank you for your attention.

