Collecting and publishing paradata on the German Internet Panel

Ulrich Krieger and Annelies Blom
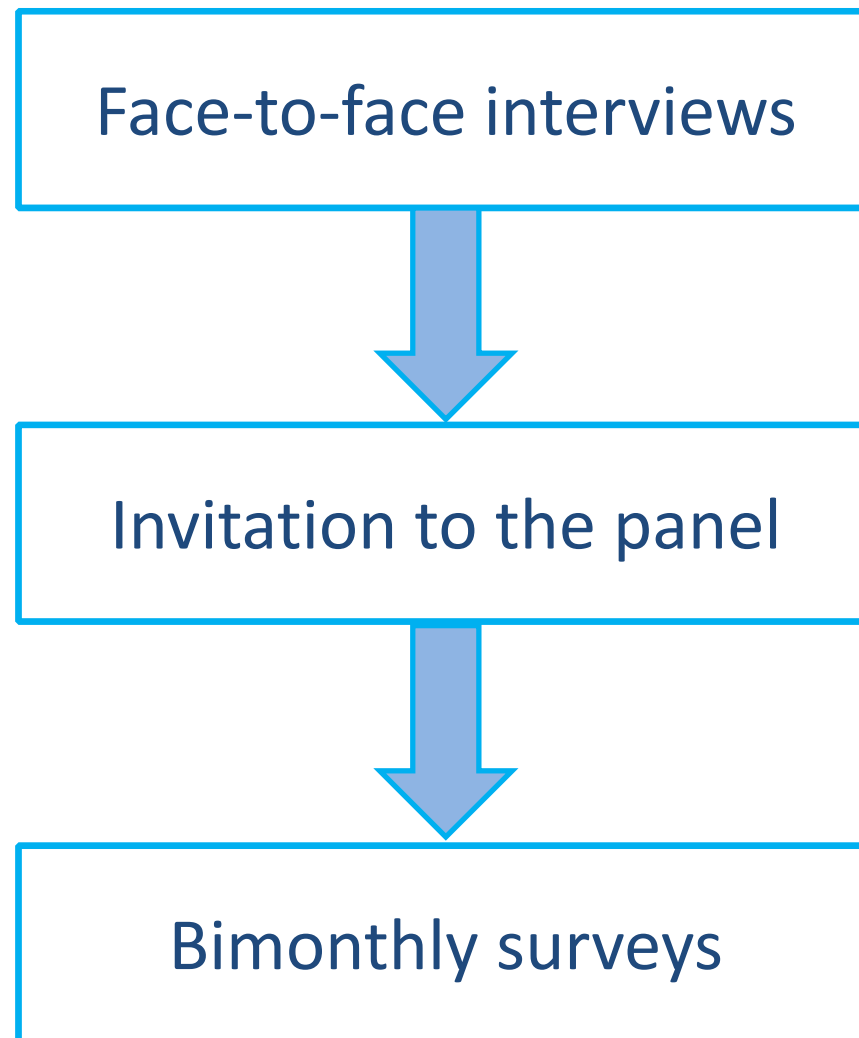
University of Mannheim

What is the GIP?

- The central data collection project of the Collaborative Research Center (SFB 884) "Political Economy of Reforms" at Mannheim University

- Interdisciplinary group of researchers: economists, political scientists and sociologists

- Funded by the German Research Foundation (DFG)

- First wave in 2012; funded until at least 12/2017

Methodology of the GIP

- Face-to-face recruited online panel via 15-20 minute CAPI interviews

- Households without internet and/or computers are equipped to participate

- Online data-collection every 2 months

- Questionnaire length approx. 20 minutes

GIP recruitment phases

```
┌─────────────────────────────────┐
│     Face-to-face interviews     │
└─────────────────────────────────┘
                 ↓
┌─────────────────────────────────┐
│      Invitation to the panel    │
└─────────────────────────────────┘
                 ↓
┌─────────────────────────────────┐
│        Bimonthly surveys        │
└─────────────────────────────────┘
```
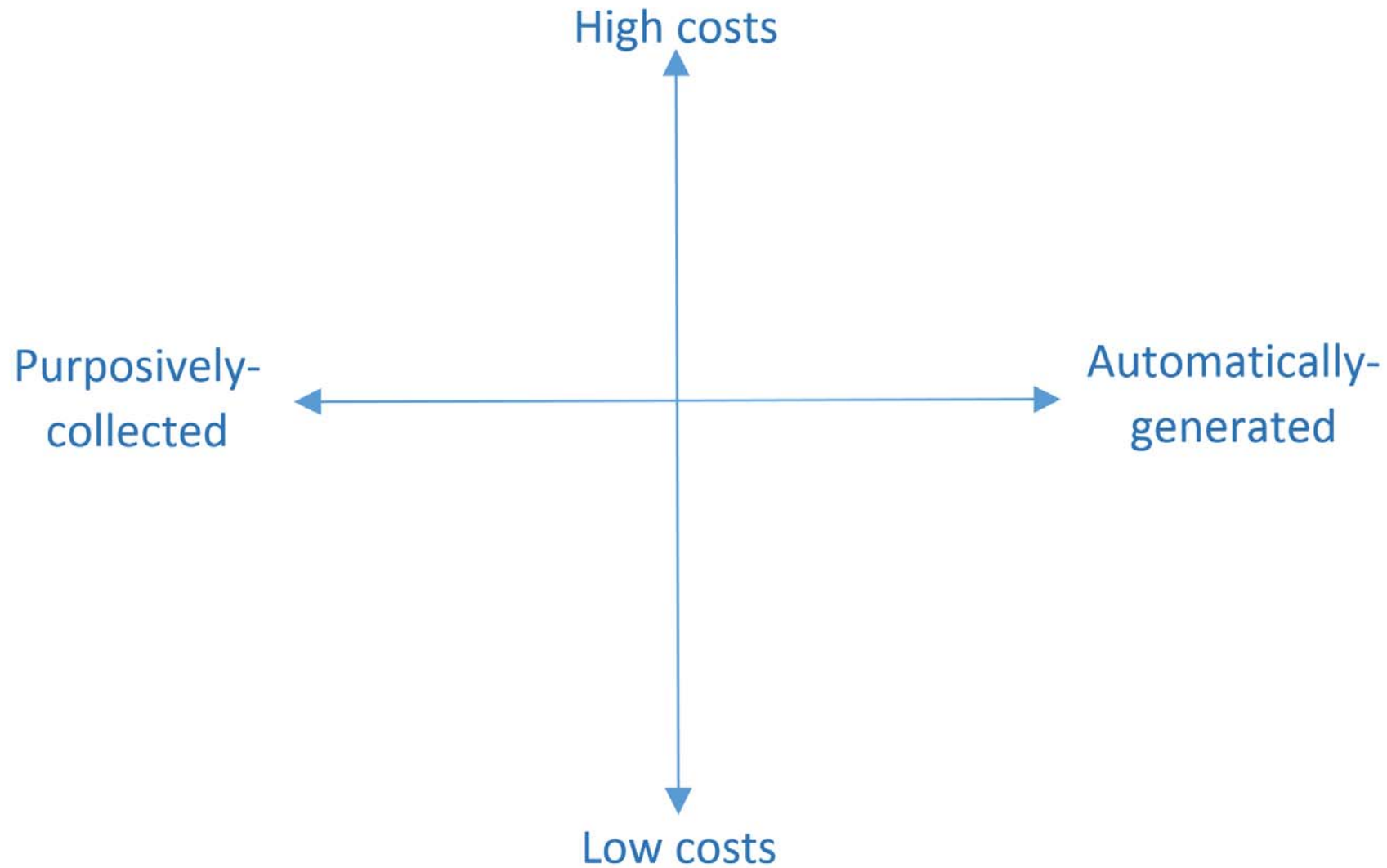
Paradata

Paradata are often considered

- automatically-generated by-products of the survey process
- available at no or very low additional costs

Examining the various types of paradata in the GIP, does this notion hold for all paradata?

Note: "Costs" also include indirect costs for ensuring data protection and ethical considerations (e.g. informed consent).
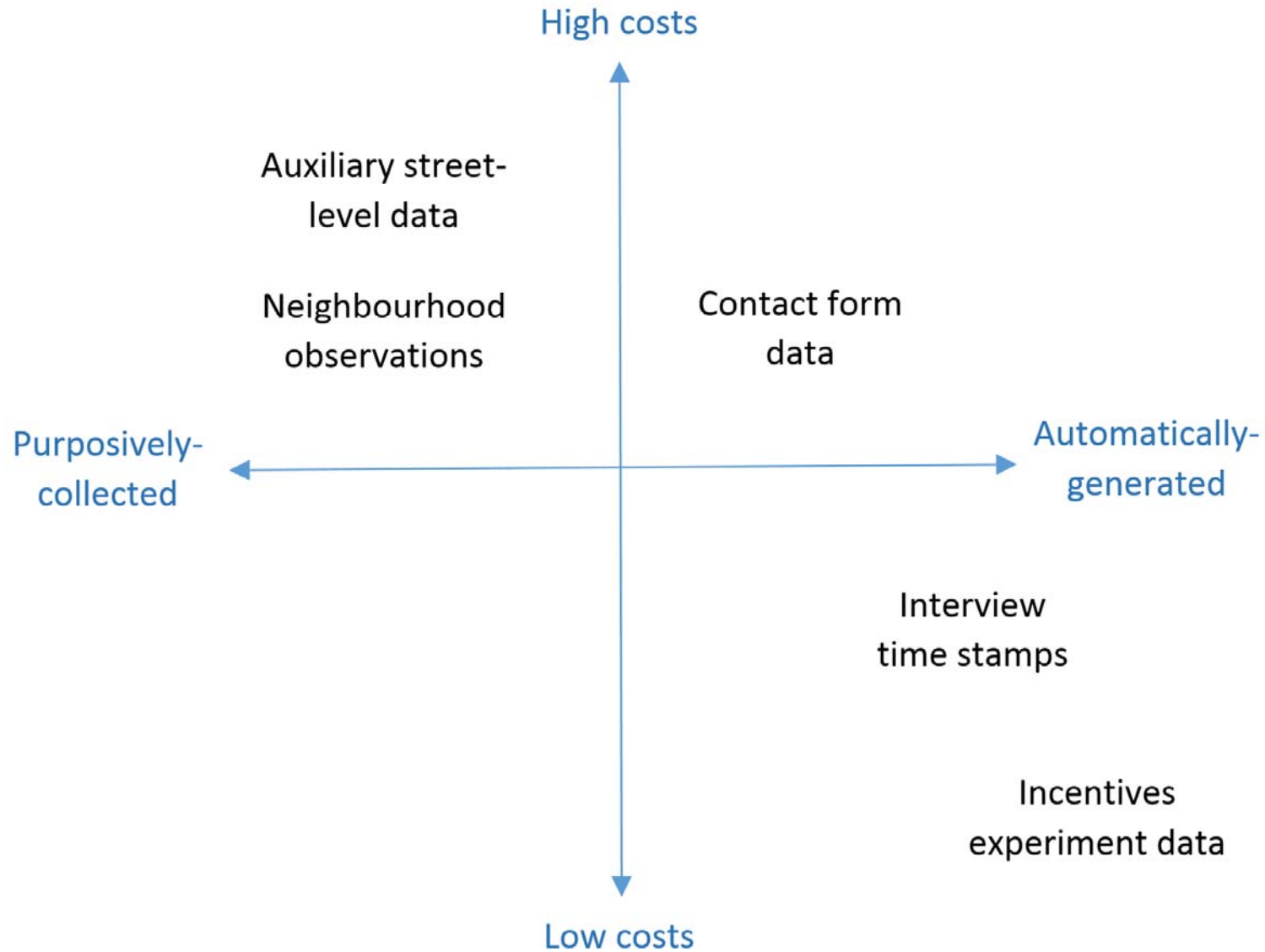
Analysis framework
for GIP paradata

Paradata during f2f recruitment

- contact form data

- neighbourhood observations

- date of CAPI interview

- auxiliary street-level data from commercial data provider

- treatment groups of incentives experiment

Paradata during f2f recruitment
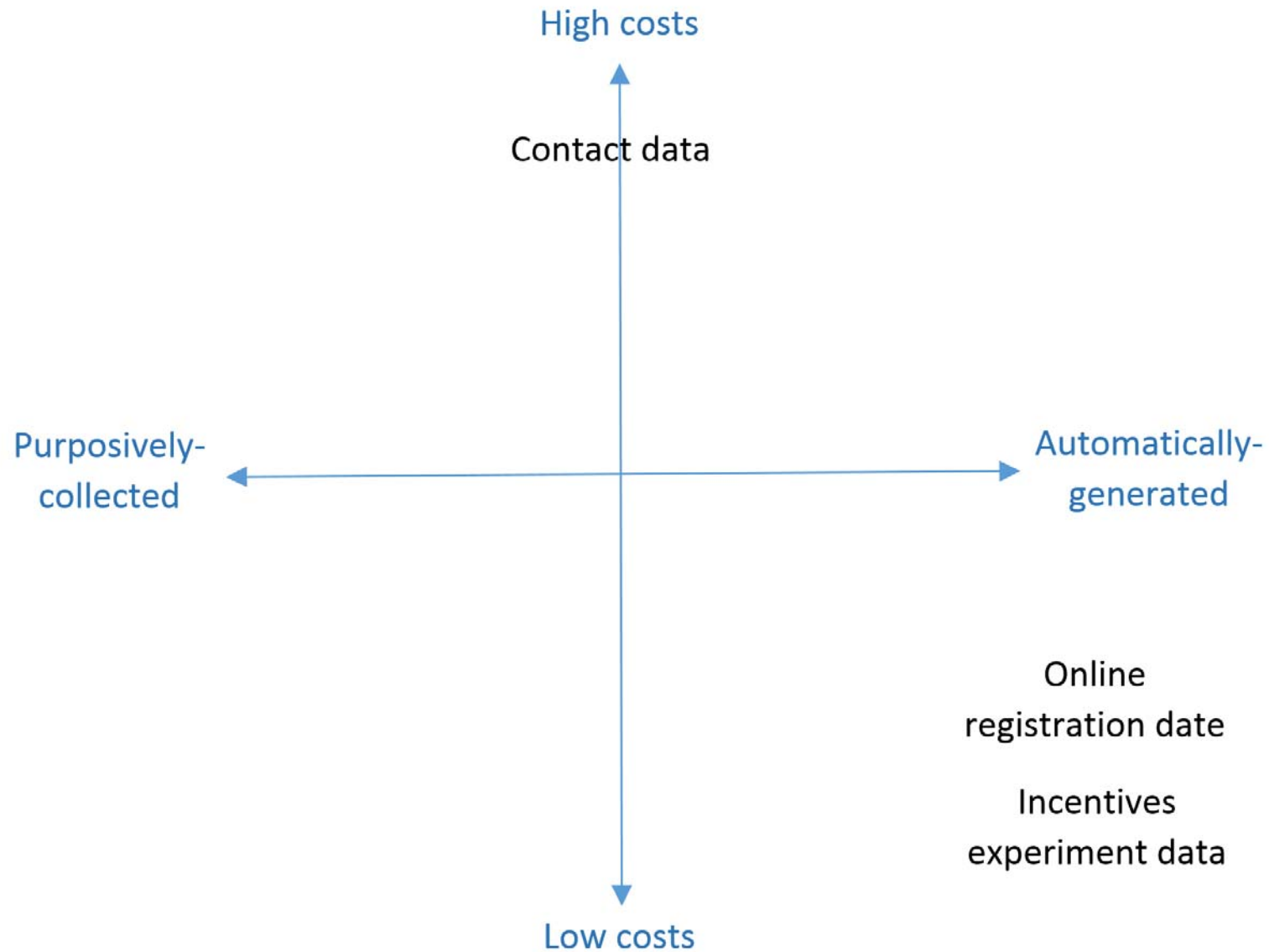

Krieger and Blom

More recruitment paradata

- more contact data, such as the date of the invitation mail-out, reminder mail-out or phone reminder

- panel registration date

- treatment groups of second incentives experiment

More recruitment paradata

High costs

Contact data

Purposively-
collected

Automatically-
generated

Online
registration date

Incentives
experiment data

Low costs

Online paradata

- client-side paradata, such as the type of operating system used by the respondent, browser types, screen resolutions, the availability of Java/JavaScript/Flash etc.

- time-stamps

- keystroke and mouse data

- paradata on nonresponse and attrition processes

- data on offliner equipment

- communication from panel members

Online paradata
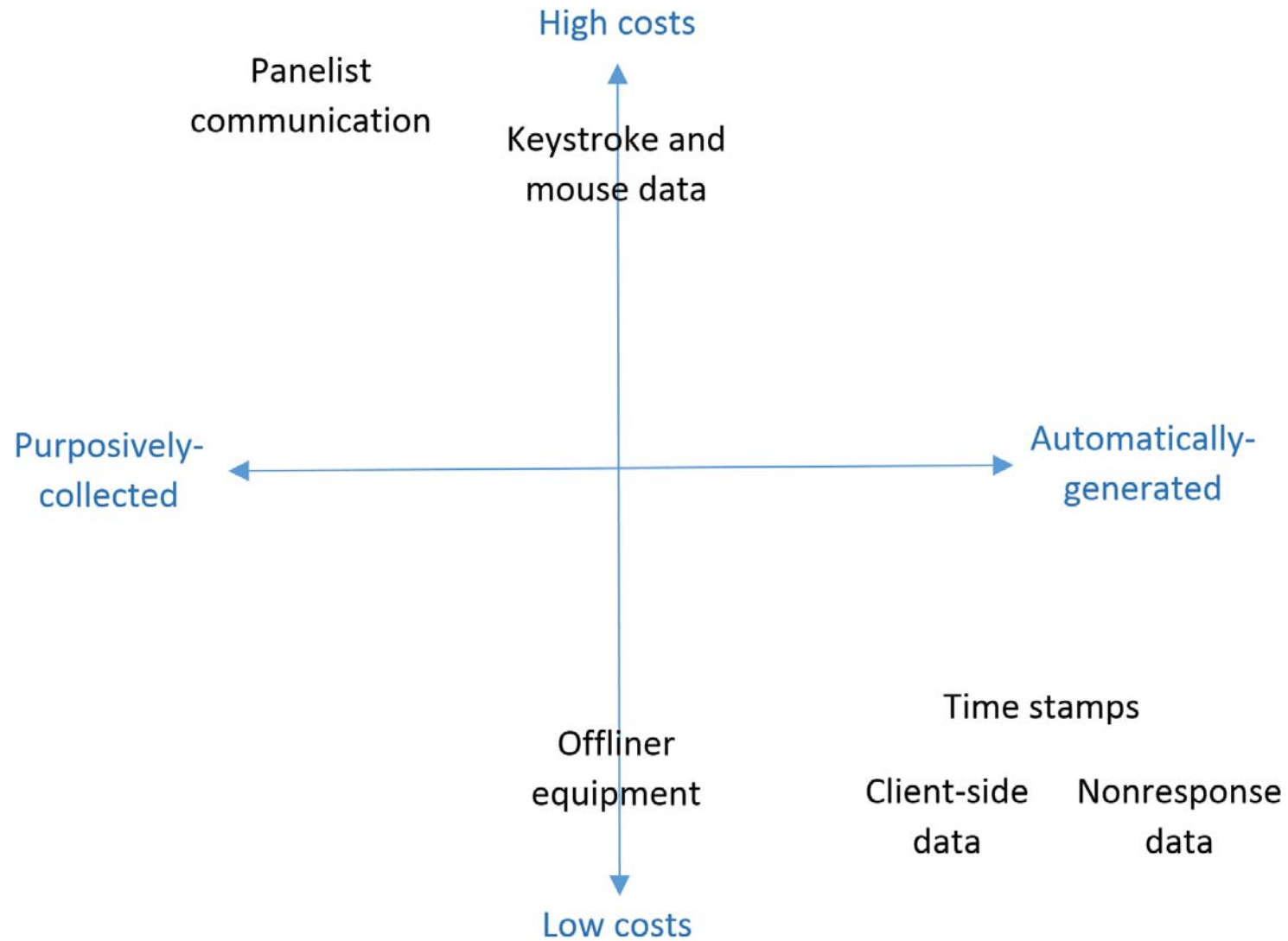
Discussion

Automatic collection:

Only about half of the GIP paradata are automatically collected .

Low costs:

For about half of the GIP paradata the costs of making them available to data analysts is high. Even low-cost paradata demand data preparation and documentation costs.

Ethical and data protection concerns:

Still many unresolved ethical and data protection concerns, e.g. informed consent:

> "Our research shows that requiring such explicit consent may reduce survey participation without adequately informing survey respondents about what paradata are and why they are being used." (Couper and Singer 2013)

Thank You!