System-to-System Data Collection in business surveys applied to an agricultural survey: small-scale pilot results

Ger Snijkers, Tim de Jong, Chris Lam, Cath van Meurs, and many CBS colleagues
22 May 2024
(EU grant: 101036345 - 2020-NL-AGRI-SISA)

Expert Meeting on Statistical Data Collection and Sources
22-24 May 2024, Geneva, Switzerland
Pre-fill

X

“Why do I still have to do this manually?”

Pilot with John Deere

Pre-filling: How to make this work?

Techno farmer has the future

Smart industries

Smart farming

© John Deere

200+ registries

Techno farmer heeft de toekomst

Fosfaatproductie in dierlijke mest per landbouwgebied in 2016
MyJohnDeere data

- Data: operations per field (event-based)
- Almost 100% overlap with data in Crop Yield Survey questionnaire

> MyJohnDeere is (potentially) a good source!

### Crop yield survey

<table>
<thead>
<tr>
<th>Grains</th>
<th>Harvested area</th>
<th>Total yield</th>
<th>Moisture content</th>
<th>Crop failure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Winter wheat</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Winter barley</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Summer barley</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rye</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oats</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tritricale</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grain corn</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Yield**

<table>
<thead>
<tr>
<th>Grains</th>
<th>Harvested area</th>
<th>Total yield</th>
<th>Moisture content</th>
</tr>
</thead>
<tbody>
<tr>
<td>Winter wheat</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Winter barley</td>
<td>harvest_area_ha: 70</td>
<td>production_t: 700</td>
<td></td>
</tr>
<tr>
<td>Summer barley</td>
<td>no_harvest_area_ha: 5</td>
<td>humidity_prnt: 5</td>
<td></td>
</tr>
<tr>
<td>Rye</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oats</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tritricale</td>
<td>harvest_area_ha: 4</td>
<td>production_t: 40</td>
<td></td>
</tr>
<tr>
<td>Grain corn</td>
<td>no_harvest_area_ha: 1</td>
<td>humidity_prnt: 5</td>
<td></td>
</tr>
</tbody>
</table>

**Crop failure**

<table>
<thead>
<tr>
<th>Grains</th>
<th>Area not harvested</th>
</tr>
</thead>
<tbody>
<tr>
<td>Winter wheat</td>
<td></td>
</tr>
<tr>
<td>Winter barley</td>
<td></td>
</tr>
<tr>
<td>Summer barley</td>
<td></td>
</tr>
<tr>
<td>Rye</td>
<td></td>
</tr>
<tr>
<td>Oats</td>
<td></td>
</tr>
<tr>
<td>Tritricale</td>
<td></td>
</tr>
<tr>
<td>Grain corn</td>
<td></td>
</tr>
</tbody>
</table>
The farmer’s completion process:
1. Comprehension
2. Data retrieval
3. Computation
4. Evaluation and reporting

Automate
Test: Sandbox + technical test

It worked!

1. Sandbox:
   - open data from John Deere
   - Virtual farm

Next:

2. Technical test:
   
   In theory the system works!
Test 3: Small-scale pilot with farmers

• How does it work in practice?
• Pre-test with 5 farmers
  o Hard to recruit!
• Pre-test results:
  o Technical issues
  o Usability issues (the farmer/user’s perspective)
  o Data quality issues
  o Perceived workload
  o Trust
  o General attitude

Assumption: the data in MyJohnDeere are correct!
The implemented system pre-tested

1. Farmer logs in to Q
The implemented system pre-tested.

The farmer’s completion process:
1. Farmer logs in to Q

Blaise Questionnaire
The implemented system pre-tested

**The farmer’s completion process:**
1. Farmer logs in to Q
2. MyJohnDeere?
The implemented system pre-tested

Blaise Questionnaire

The farmer’s completion process:
1. Farmer logs in to Q
2. MyJohnDeere?

**First design**

**Usability issues:**
- The “John Deere” button was not recognized as a button
- How to use the “Back” and “Next” buttons?
The implemented system

Revised design

Blaise Questionnaire

The farmer’s completion process:
1. Farmer logs in to Q
2. MyJohnDeere?
The implemented system pre-tested

The farmer’s completion process:
1. Farmer logs in to Q
2. MyJohnDeere?
3. Authentication
   Import data?
The implemented system pre-tested

In practice: more complex process

The farmer's completion process:
1. Farmer logs in to Q
2. MyJohnDeere?
3. Authentication

Technical issues:
• Authentication did not work properly: two-step procedure
• Instable systems
• Unreliable communication between systems
• Not all retrieved data were shown in the questionnaire

Blaise Questionnaire

1, 2

Log in to MyJohnDeere
Authentication Microservice

In practice:
more complex
process

MyJohnDeere cloud

Crop Yield Survey 2022
Due date: 1-1-2023

Authentication
✓ The authentication was successful! You can now continue to fill out the questionnaire.

Do you agree to use the retrieved data fill in questions in the questionnaire?

- Yes
- No

Back Next
The implemented system pre-tested

**The farmer’s completion process:**

1. Farmer logs in to Q
2. MyJohnDeere?
3. Authentication
4. Blaise Q <-> Microservice <-> John Deere
5. Data are pre-filled
The implemented system pre-tested

The farmer's completion process:
1. Farmer logs in to Q
2. MyJohnDeere
3. Authentication
4. Blaise Q <-> Microservice <-> John Deere
5. Data are pre-filled

Calculated answers by Data Collection Microservice (JSON output)

Farmers didn’t recognise these totals

**Crop Yield Survey 2022**

Due date: 1-1-2023

Ophalen data John Deere
Result calculated crops Standard

<table>
<thead>
<tr>
<th>Crop</th>
<th>harvest_area_ha</th>
<th>production_t</th>
<th>no_harvest_area_ha</th>
<th>humidity_prcnt</th>
</tr>
</thead>
<tbody>
<tr>
<td>winterwhear</td>
<td>70.00</td>
<td>700.0</td>
<td>5.0</td>
<td>5.0</td>
</tr>
<tr>
<td>zomerwhear</td>
<td>harvest_area_ha</td>
<td>production_t</td>
<td>no_harvest_area_ha</td>
<td>humidity_prcnt</td>
</tr>
<tr>
<td>wintergerst</td>
<td>harvest_area_ha</td>
<td>production_t</td>
<td>no_harvest_area_ha</td>
<td>humidity_prcnt</td>
</tr>
<tr>
<td>zomergerst</td>
<td>harvest_area_ha</td>
<td>production_t</td>
<td>no_harvest_area_ha</td>
<td>humidity_prcnt</td>
</tr>
<tr>
<td>rogers</td>
<td>harvest_area_ha</td>
<td>production_t</td>
<td>no_harvest_area_ha</td>
<td>humidity_prcnt</td>
</tr>
</tbody>
</table>

**calculated_crops:**

- **winterwhear**
  - harvest_area_ha: 70
  - production_t: 700
  - no_harvest_area_ha: 5
  - humidity_prcnt: 5

- **rye**
  - harvest_area_ha: 4
  - production_t: 40
  - no_harvest_area_ha: 1
  - humidity_prcnt: 5

- **corn**
  - {}...

- **seed_onion**
  - {}...
Pre-test results

• Data quality issues:
  o Farmers indicated that data in “MyJohnDeere” may not be correct:
    - not calibrated (sensor calibration)
    - data in MyJohnDeere cannot be edited
    - MyJohnDeere is not designed to be a Farm Management Information System; primary purpose is for machine maintenance
    - Farmers used their FMIS to check the data (Dacom & AgroVision)
  o Missing data:
    - Crops harvested with machines not connected to MyJohnDeere:
      JohnDeere tractors, other brands
    - Crops harvested by contractors
  o Unit issues:
    - Data from neighbours: helping out
  o Selectivity:
    - Market share (small; FMIS: 50% of arable farmers) and take-up rate

Assumption: the data in MyJohnDeere are correct!
The implemented system pre-tested

The farmer’s completion process:
1. Farmer logs in to Q
2. MyJohnDeere?
3. Authentication
4. Blaise Q <-> Microservice <-> John Deere
5. Data are pre-filled
6. Check, edit, and add
7. Submit
### Winter wheat estimates

**Production data from 2021**

*Please complete all answer fields and make an estimation if necessary. Enter a 0 in the fields that are not applicable.*

*Round up area questions to two decimal places and other questions to one decimal place.*

<table>
<thead>
<tr>
<th>Winter wheat estimates</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Harvested area of winter wheat</strong></td>
<td>70.00 ha</td>
</tr>
<tr>
<td><strong>Total winter wheat</strong></td>
<td>700.00 ton</td>
</tr>
<tr>
<td><strong>Yield per hectare</strong></td>
<td>10.00 ton/ha</td>
</tr>
<tr>
<td><strong>Percentage moisture content</strong></td>
<td>5.0 %</td>
</tr>
<tr>
<td><strong>Area not harvested</strong></td>
<td>5.00 ha</td>
</tr>
</tbody>
</table>

**3. Authentication**

Import data?

**6. Check, edit,**

**7. Submit**
Pre-test results

• Perceived workload:
  o “This doesn’t make it easier.”
  o “This doesn’t reduce the time I need compared to completing the questionnaire in the usual way.”

• Trust:
  o Trust in the government
    - Farmers don’t trust the government with their data: data are NOT shared
  o Trust in the system
    - Safe and secure data communication
    - Farmers are unaware of safety measures being taken: penetration test (to find leaks, prevent hacking)
Pre-test results

General conclusions:

• General attitude:
  o These farmers were positive about the S2S approach
  o It could work, but improvements are needed to make it work in practice

• Selective group of farmers:
  o Innovative farmers
  o Positive attitude towards data and innovations
  o They are the early adopters!

• “Use FMIS systems instead”: better source to connect to!
Conclusions

• Go/No-Go decision: not implemented in the Crop Yield Survey
  o Too many issue: the risks of failure were too high. This operationalisation was not efficient for farmers
  o Low market share and low take-up rate
  o Production issues for this operationalisation of the methodology: maintainability, scalability, and costs were not met, compared to the assets
  o No time / resources for improvements
• Still: we have a working proof-of-concept

This was the goal of this project
Next step

Farm Management Information Systems (FMIS):

• Two most-used systems in Netherlands:
  
  o AgroVision
  
  o DACOM

• 35-55% of farmers:
  
  o Crop Yield Survey: AgroVision 21%, Dacom 5%, other 7% of farmers
  
  o Annual Agricultural Counts: 56% of all farmers use a FMIS
  
  o 45% of fields with the largest crops is registered in AgroVision

• Next project: connect to these systems

> positive business case!
What do you think ...
- Is this a feasible data collection method?
- Experiences?

Ger Snijkers: g.snijkers@cbs.nl