DEUTSCHE BUNDESBANK EUROSYSTEM

Introduction to and application of SDC rules using self-developed tools

J. Blaschke, M. Gomolka, C. Hirsch, S. Seltmann and H. Stahl (Deutsche Bundesbank)

11th Biennial IFC Conference on "Post-pandemic landscape for central bank statistics"

Session 2 – Microdata disclosure control: a practical perspective

25 August 2022

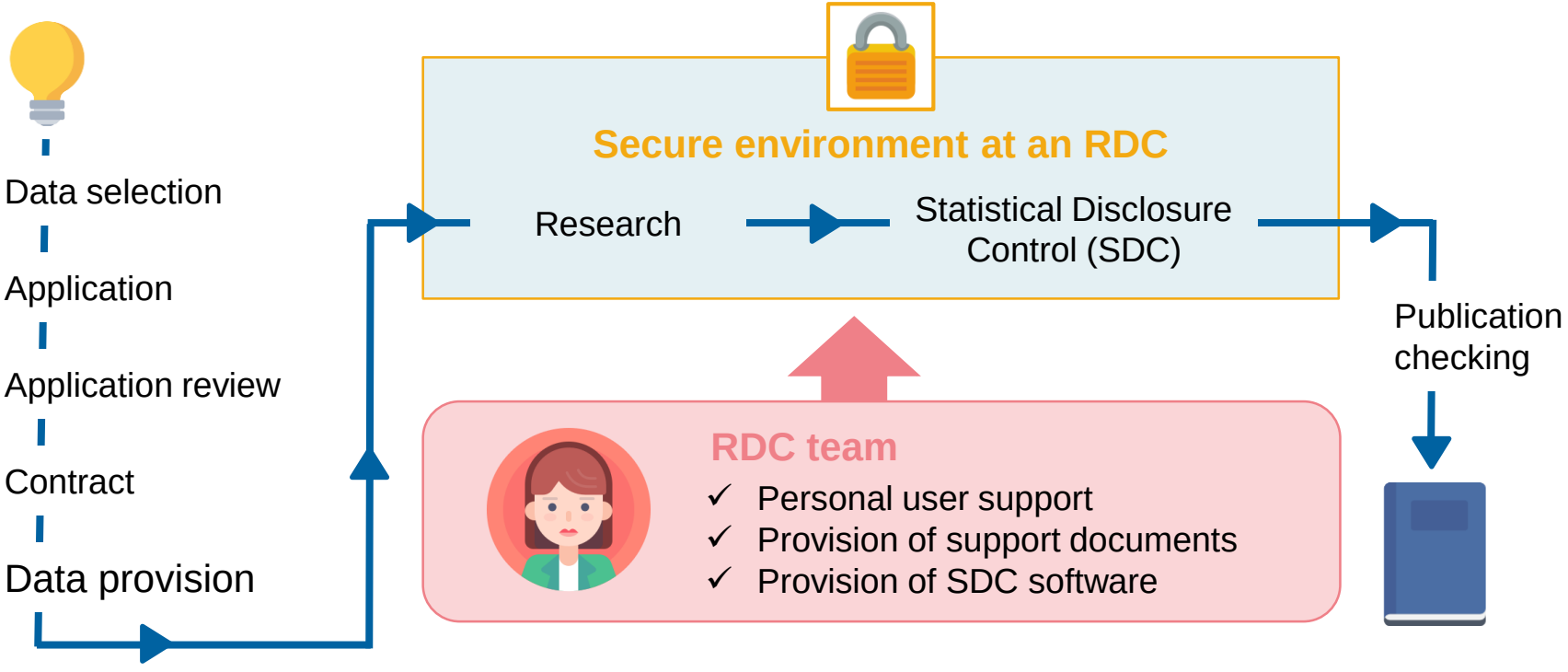
The views expressed here do not necessarily reflect the opinion of the Deutsche Bundesbank, the INEXDA network or the Eurosystem.

Research Data and Service Centre

Introduction to and application of SDC rules using self-developed tools

1. A brief introduction to the work of Research Data Centres (RDCs) (1|2)

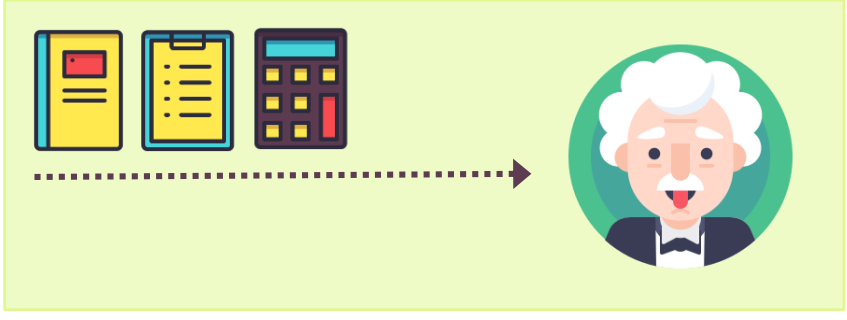
RDCs provide secure on-site access to confidential micro data for scientific research



Research Data and Service Centre

Introduction to and application of SDC rules using self-developed tools
1. A brief introduction to the work of Research Data Centres (RDCs) (2|2)

Provision of support materials (e.g. documents, software) can potentially lead to an information overflow



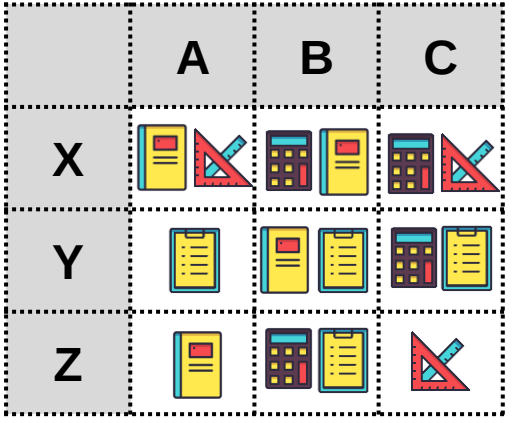
Scenario 1 - Little user support needed

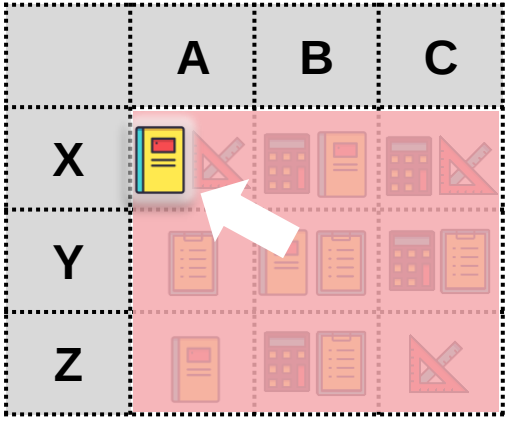
- Small number of datasets
- Easy data structure
- Small dataset size
- Homogeneous legal framework → Similar rules for data access and SDC

Scenario 2 - Much user support needed

- Large number of datasets
- Complex data structure (e.g. multiple IDs, missing IDs)
- Large dataset size
- Heterogeneous legal framework → Dataset-specific rules for data access and SDC

***Possible solution for scenario 2**: Pre-sorting information and automating as far as possible*





Case A

- ☐ No structure of material
- ☐ No Customized filtering

Case B

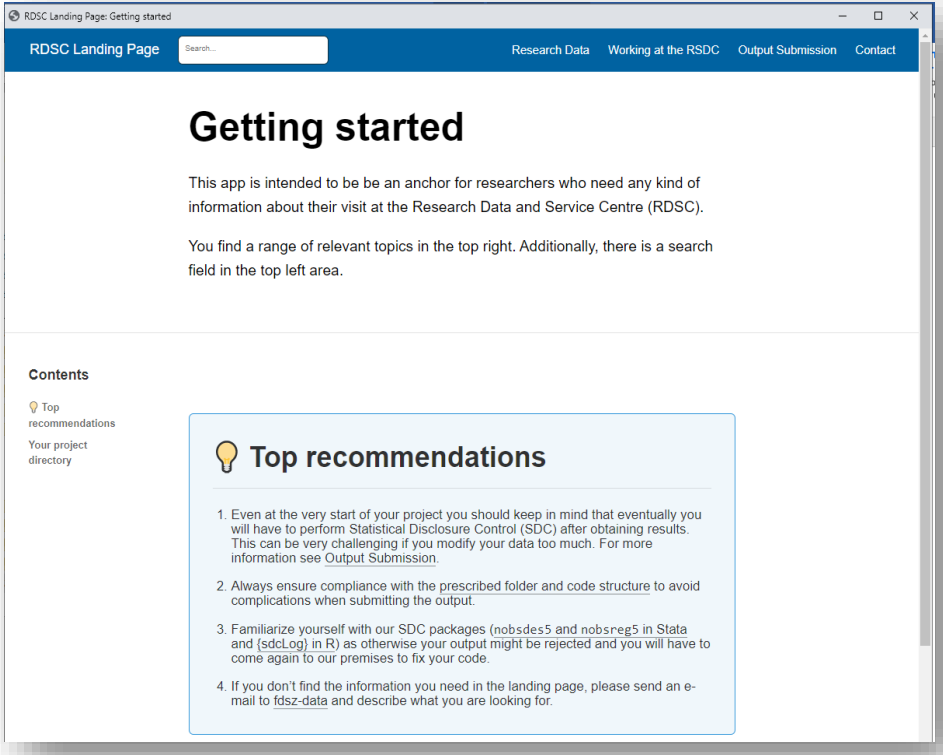
- ✓ Structured information (e.g. by researcher's characteristics and purpose of the document)
- ☐ No Customized filtering

Case C

- ✓ Structured information (e.g. by researcher's characteristics and purpose of the document)
- ✓ Customized filtering

Research Data and Service Centre

***The Project start**: Researchers would like to get an overview of all available information*



Features

- Structured **overview** of all available resources (e.g. documentation, software)
- Show **relations** between similar information across documents → Point researchers to relevant information they were not actually searching

➡ *Structured information*

Technical set-up

- NW.js application, which is built using Rmarkdown and the R package {distill}
- Advantages: Easy to modify, possibility to open links to local files

Research Data and Service Centre

During the research: *Researchers need information that applies directly to their analysis*

nobsdes5
nobsreg5

{sdcLog}

- Both packages are freely available on the RDSC's website.
- {sdcLog} is also available on CRAN https://cran.r-project.org/package=sdcLog

Features

Researchers can use commands after generating a result (e.g. descriptive or regression table) and get immediate feedback, if the table will pass SDC or not

- ➡ No need to program checks themselves
- ➡ Prerequisite: Researchers need to correctly apply the commands

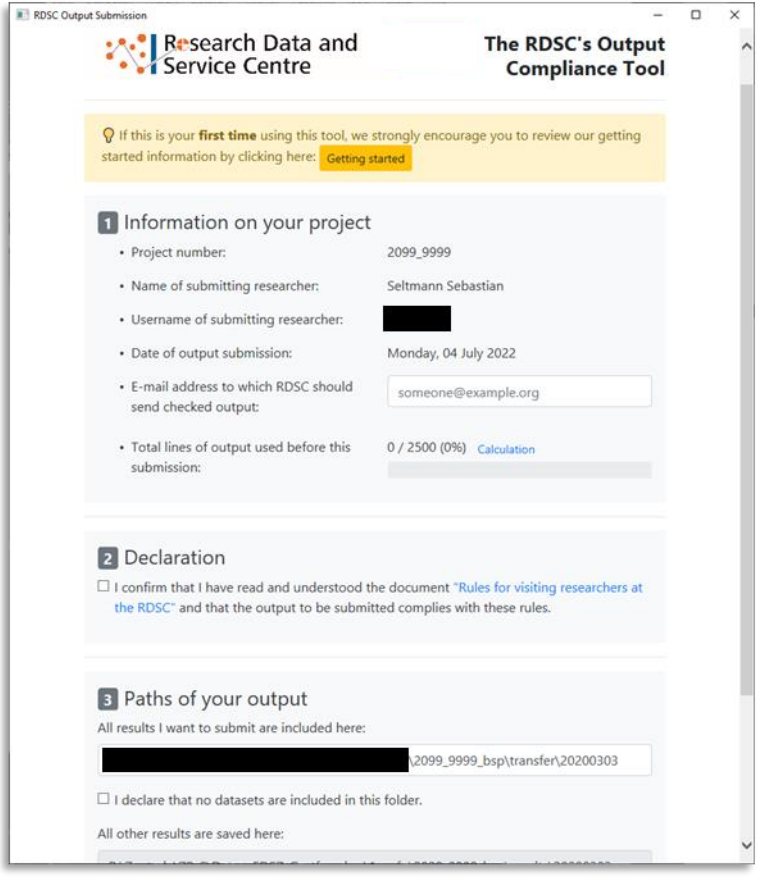
Example from nobsdes5:

```
. nobsdes5 id x, by(year) notab

D I S C L O S U R E problem:
Share of largest two IDs > 85%
Smallest number of distinct IDs (id) of variable x for year: too small
```

Introduction to and application of SDC rules using self-developed tools

3. Practical examples for self-developed BBk tools - RDSC Output Submitter (1|2)



Finalizing the project: *Researchers need to ensure compliance with SDC rules*

Features

- Automated checks for selected rules (e.g. file format, folder structure)
- Warnings for potential rule breaches
- Automated count of remaining lines of output

➡ *Customized filtering of information*

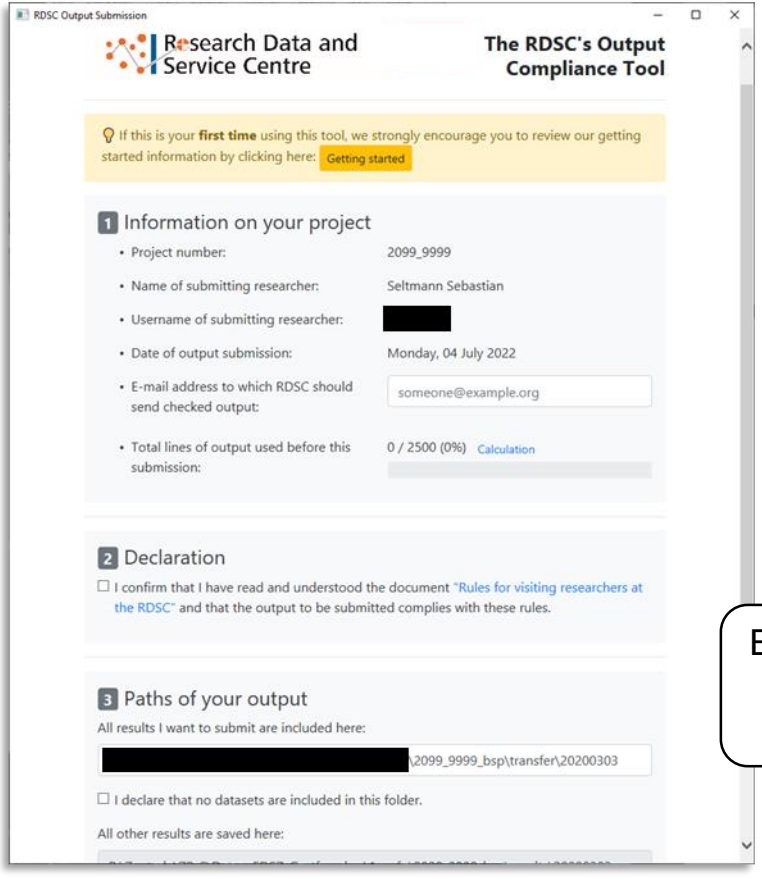
Technical set-up

- Web-technologies (javascript, html, css, node)
- Runs locally, allowing access to research project code, logs and output files

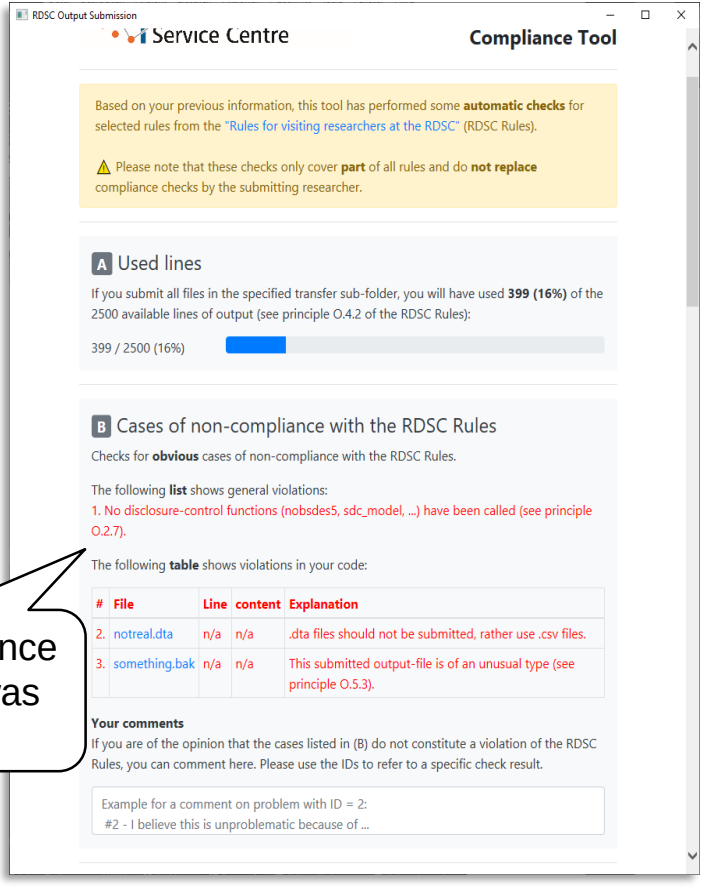
Research Data and Service Centre

Introduction to and application of SDC rules using self-developed tools

3. Practical examples for self-developed BBk tools - RDSC Output Submitter (2|2)

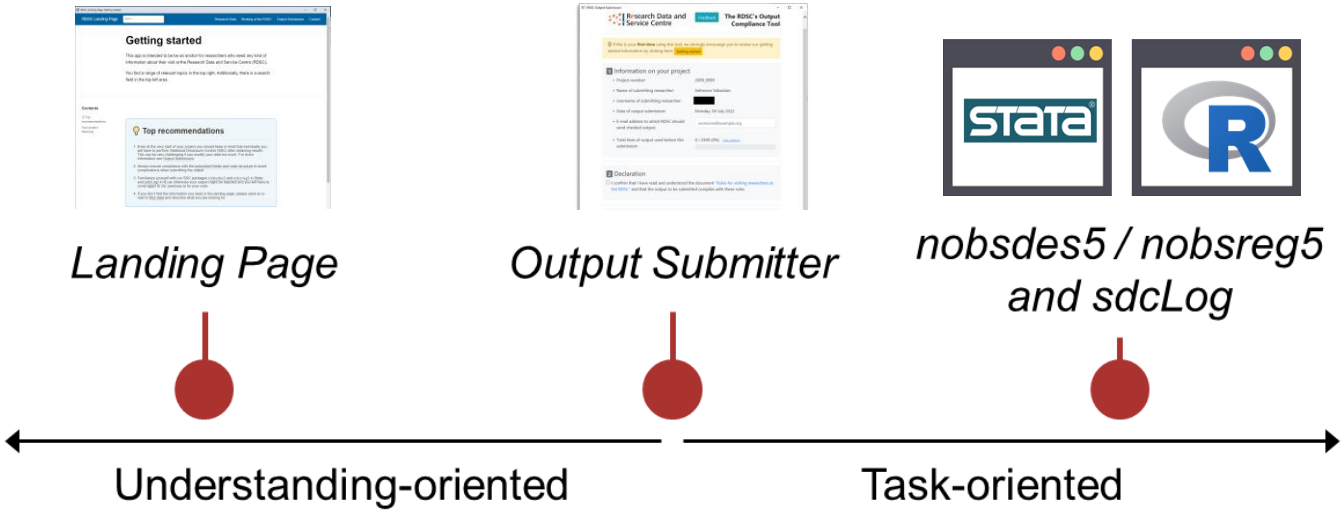


Example



Explicit reference to rule that was breached.

Research Data and Service Centre



Future areas of improvement

Move from semi-automated to **fully automated tools**. While this would clearly benefit researchers as they get more help in performing SDC it also complicates the development of these tools putting some constraints on RDCs to do this.

Thank you

Jannick.blaschke@bundesbank.de

Website: www.bundesbank.de/rdsc

Contact: fdsz@bundesbank.de

Research Data and Service Centre