

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

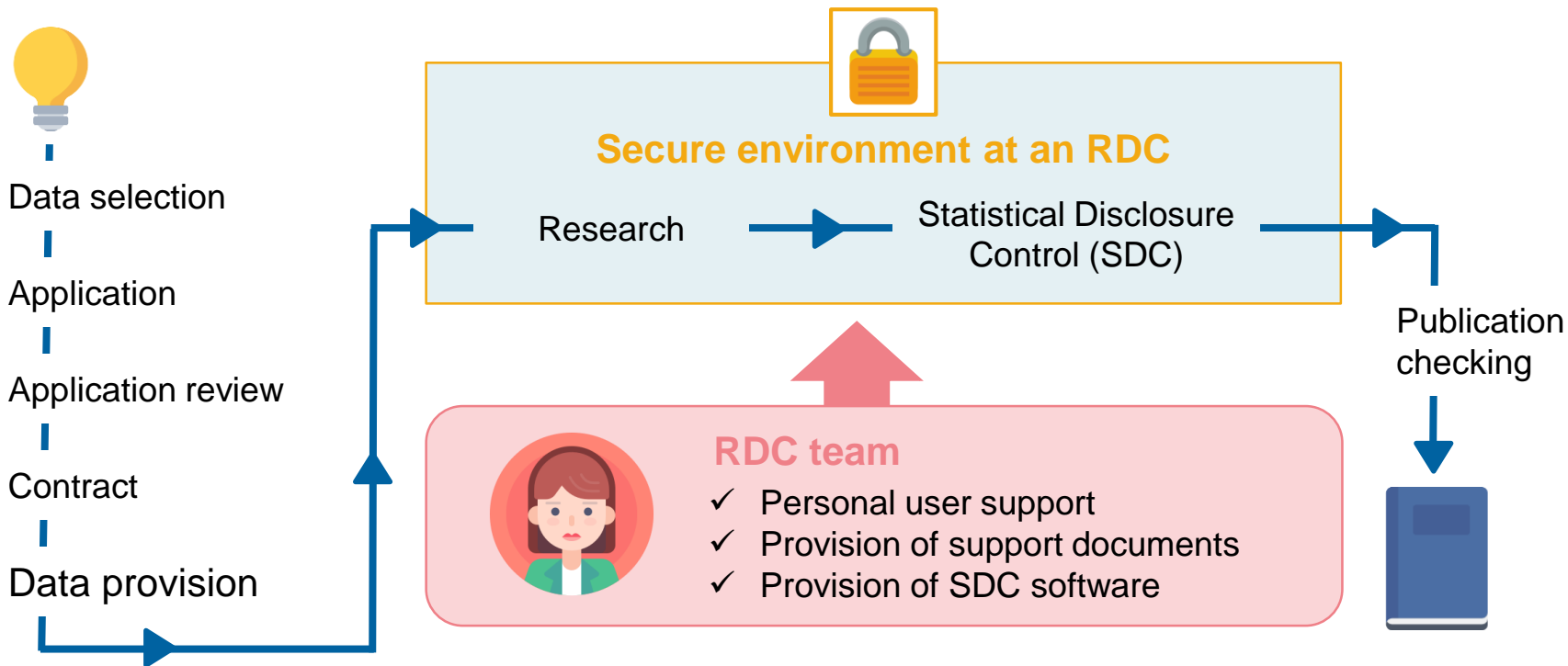
25 August 2022

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

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)

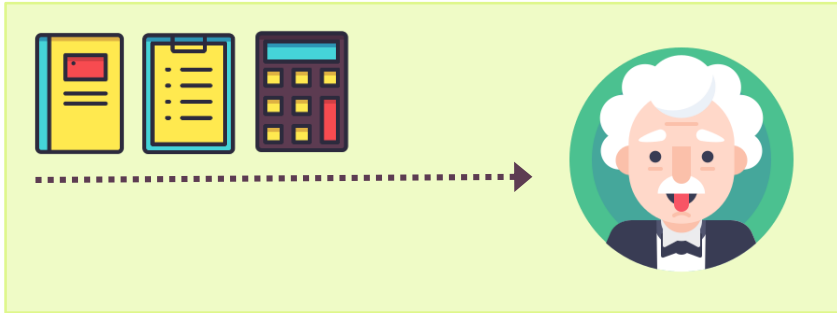
RDCs provide secure on-site access to confidential micro data for scientific research



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

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
















2. How tools support researchers in understanding and applying SDC rules

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


















Case A

- ☐ No structure of material
- ☐ No Customized filtering

	A	B	C
X	 	 	 
Y		 	 
Z		 	

Case B

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

	A	B	C
X	 	 	 
Y		 	 
Z		 	

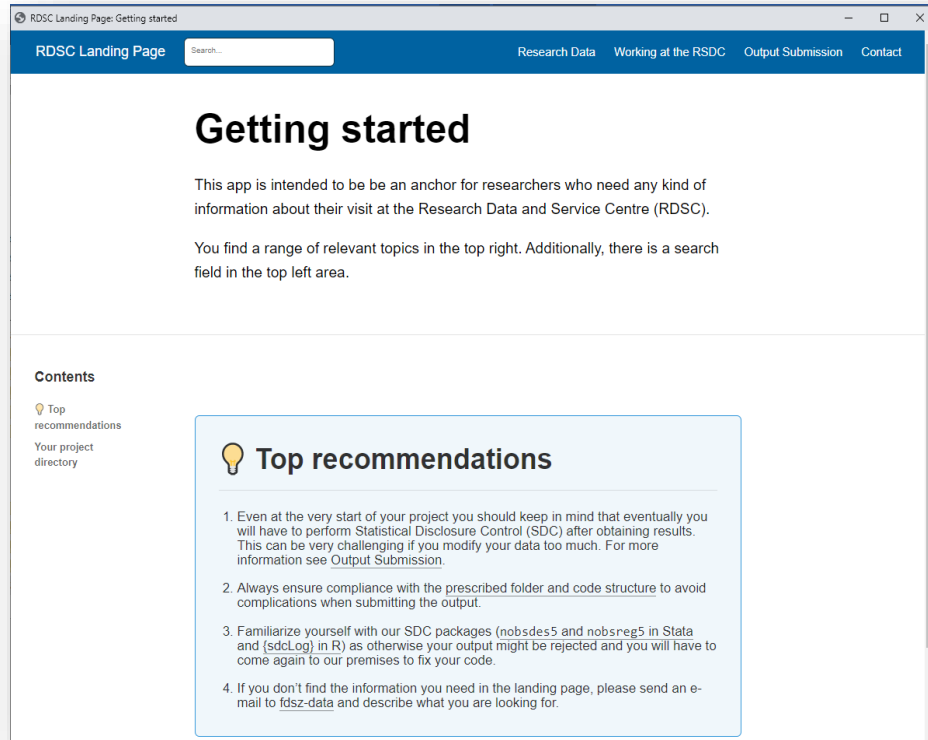
Case C

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

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

3. Practical examples for self-developed BBk tools - [RDSC Landing Page](#)

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

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

3. Practical examples for self-developed BBk tools - **SDC tools**

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)

The screenshot shows the 'RDSC Output Submission' web application. The header includes the Research Data and Service Centre logo and the title 'The RDSC's Output Compliance Tool'. A yellow banner at the top contains a message: 'If this is your first time using this tool, we strongly encourage you to review our getting started information by clicking here: [Getting started](#)'. The main form is divided into three sections: 1. 'Information on your project' with fields for Project number (2099_9999), Name of submitting researcher (Seltmann Sebastian), Username of submitting researcher (redacted), Date of output submission (Monday, 04 July 2022), E-mail address to which RDSC should send checked output (someone@example.org), and Total lines of output used before this submission (0 / 2500 (0%) with a 'Calculation' link). 2. 'Declaration' with a checkbox for confirming that the user has read and understood the 'Rules for visiting researchers at the RDSC' and that the output complies with these rules. 3. 'Paths of your output' with a text area for 'All results I want to submit are included here:' (containing a redacted path and '2099_9999_bsp\transfer\20200303') and a checkbox for 'I declare that no datasets are included in this folder.' Below this is a text area for 'All other results are saved here:'.

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

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

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

RDSC Output Submission

Research Data and Service Centre

The RDSC's Output Compliance Tool

If this is your **first time** using this tool, we strongly encourage you to review our getting started information by clicking here: [Getting started](#)

1 Information on your project

- Project number: 2099_9999
- Name of submitting researcher: Seltmann Sebastian
- Username of submitting researcher: [REDACTED]
- Date of output submission: Monday, 04 July 2022
- E-mail address to which RDSC should send checked output:
- Total lines of output used before this submission: 0 / 2500 (0%) [Calculation](#)

2 Declaration

☐ I confirm that I have read and understood the document "Rules for visiting researchers at the RDSC" and that the output to be submitted complies with these rules.

3 Paths of your output

All results I want to submit are included here:

☐ I declare that no datasets are included in this folder.

All other results are saved here:



Example

Explicit reference to rule that was breached.

RDSC Output Submission

Service Centre

Compliance Tool

Based on your previous information, this tool has performed some **automatic checks** for selected rules from the "Rules for visiting researchers at the RDSC" (RDSC Rules).

⚠ Please note that these checks only cover **part** of all rules and do **not replace** compliance checks by the submitting researcher.

A Used lines

If you submit all files in the specified transfer sub-folder, you will have used **399 (16%)** of the 2500 available lines of output (see principle O.4.2 of the RDSC Rules):

399 / 2500 (16%)

B Cases of non-compliance with the RDSC Rules

Checks for **obvious** cases of non-compliance with the RDSC Rules.

The following **list** shows general violations:

- No disclosure-control functions (nobsdes5, sdc_model, ...) have been called (see principle O.2.7).

The following **table** shows violations in your code:

#	File	Line	content	Explanation
2.	notreal.dta	n/a	n/a	.dta files should not be submitted, rather use .csv files.
3.	something.bak	n/a	n/a	This submitted output-file is of an unusual type (see principle O.5.3).

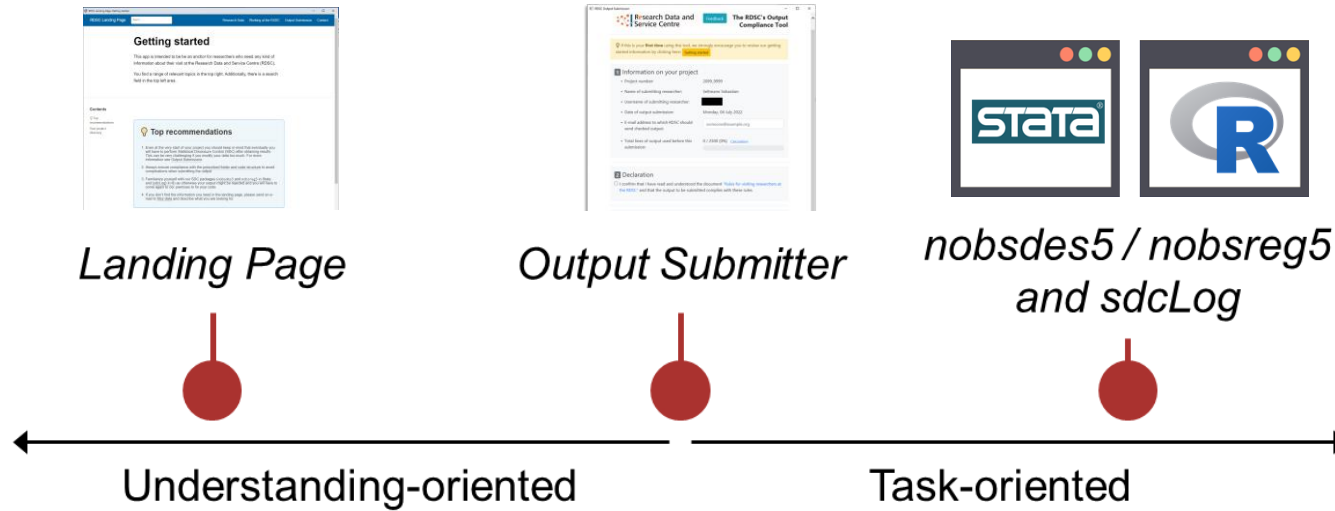
Your comments

If you are of the opinion that the cases listed in (B) do not constitute a violation of the RDSC Rules, you can comment here. Please use the IDs to refer to a specific check result.

Example for a comment on problem with ID = 2:
#2 - I believe this is unproblematic because of ...

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

4. Conclusion and discussion



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