«The reciprocal conversion of environmental data for customer information support»

L.O. Peretyatko, A.M. Sterin, Y.R. Koftan

All-Russia Research Institute of Hydrometeorological Information – World Data Center

(RIHMI-WDC)

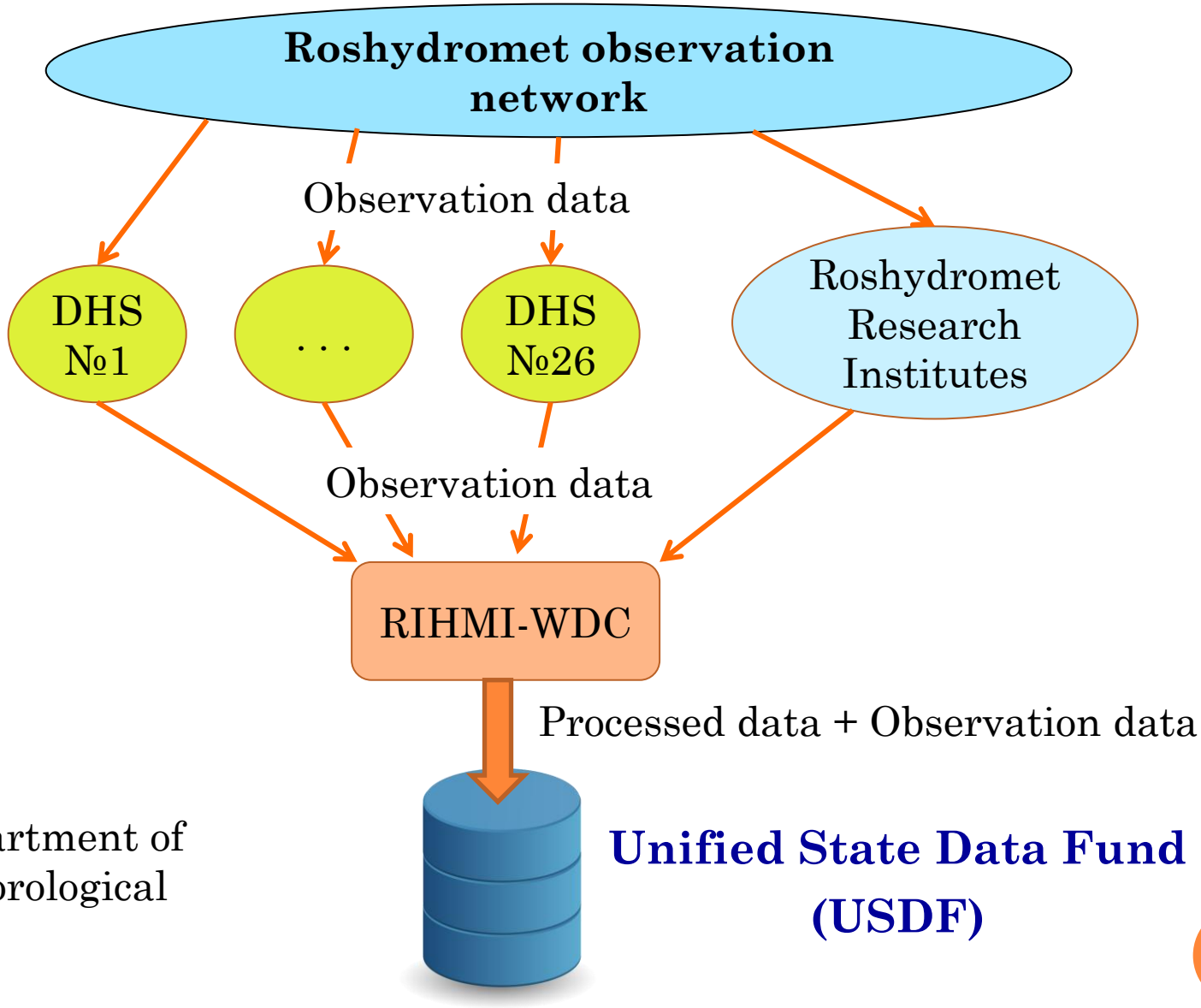
6,Korolyov St.,Obninsk,Kaluga Region,249035 Russian Federation

E-mail: Leon0121@yandex.ru
Web site: http://www.meteo.ru

Report Structure

- Roshydromet & Unified State Data Fund (USDF).
- USDF data.
- Main objectives.
- DDL as USDF data storage format with examples.
- The first version of the reciprocal data conversion system.
- Description of some algorithms and subsystems.
- Current results & Conclusion.

ROSHYDROMET & UNIFIED STATE DATA FUND



DHS - Department of Hydrometeorological Service

USDF DATA

- USDF data can be considered as Big Data, because they meet the characteristic "3V" – *volume*, *velocity*, *variety*.

- For long-term storage with the preservation of the hierarchical structure of environmental data obtained from observation networks, a specialized format of data – DDL (Hydrometeorological Data Description Language) was developed at RIHMI-WDC.

- The data in the DDL format is a combination of files – a file with a description of the data structure, and one or more files directly with the data.

Main objectives

Due to the fact the data of primary observations are of the greatest interest (can be considered as Big Data), taking into account their specifics, it is necessary to create:

1) A single technology for all types of data storage, verification (completeness and reliability of data) and provision of USDF data to consumers in the format necessary for solving their problems.

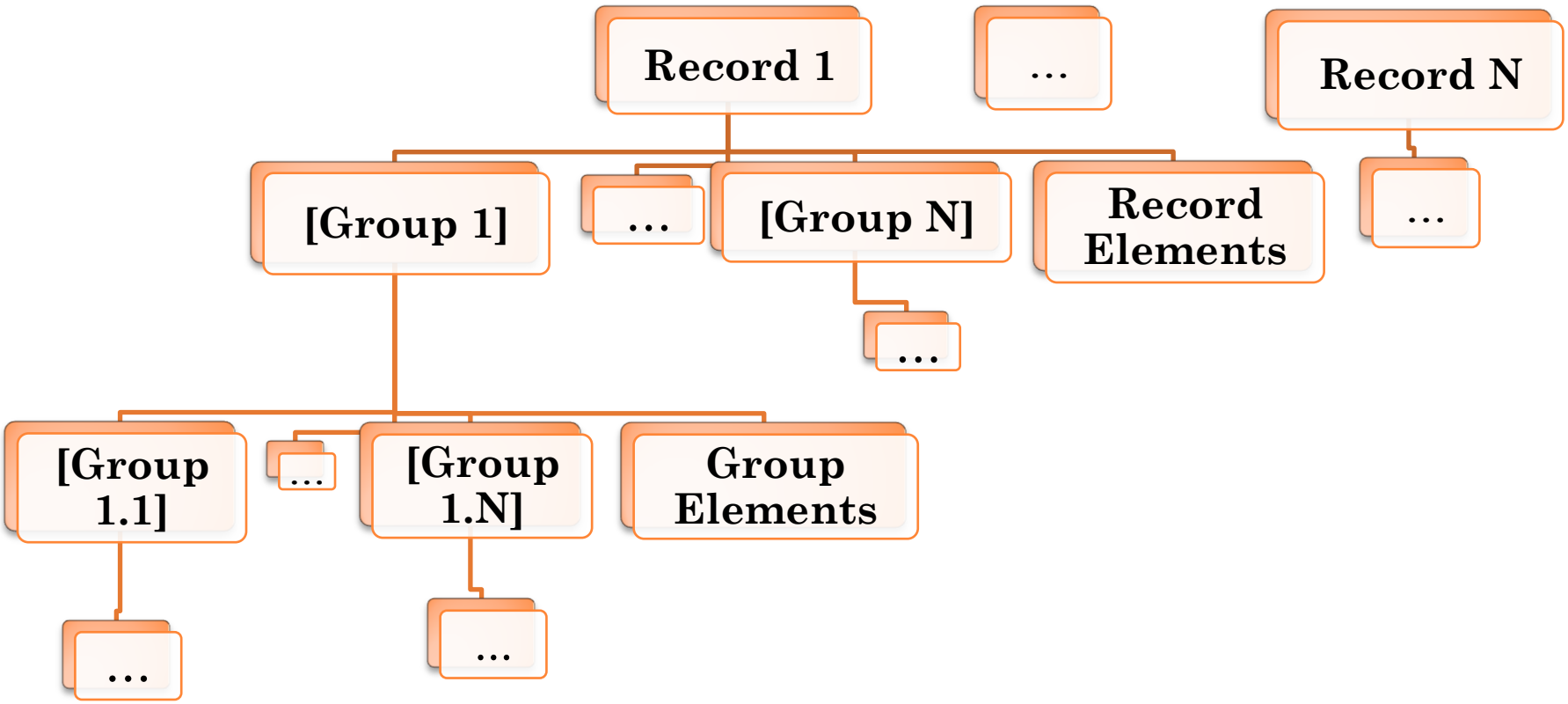
2) Technology for the formation and storage of meta descriptions (FSMD), describing the content of files and archives (file collections) of data. The meta description is information about the internal content and data state of each file.

3) Technology of reciprocal conversion of USDF data (from HDDL format to other formats widely used by consumers).

This report is dedicated to the system for reciprocal data conversion, with control over the adequacy of the conversion performed.

To be more precise - the first version of it.

General hierarchical structure of the USDF data in the DDL format



Part of DDL description of meteorological data

1) Description of the data header

```
RECORDS;
  LNG ДЛЗАП  B(2)  PC(4);
  MIT НУЛИ  B(2)  PC(4);
  KEY(I) ГОД      B(2) PC(4);   // Год
  KEY(I) МЕСЯЦ    B(1) PC(2);   // Месяц
  KEY(U) СТАНЦИЯ  B(4) PC(7);
 MRC(I)  ТИПЗАП   B(1)  PC(1);  //  Тип
записи (1-3)
```

2) Part of the record CONST description

```
 RBODY(1) CONST ; // Пасп-ые данные
  MIT НАИМЕНСТ A(20) PA(20) NA;
  MIT КООРДНОМ B(4) PC(7) NA; //
 Коорд. ном. станц
  MIT НОМУПРАВ B(1) PC(2) NA;  //
 Номер УГМС
  MIT НОМЧАСП  B(1) PC(2) NA;  //
 Номер час. пояса
  MIT ПРГЕОРАС B(1) PC(1);
  MIT КОЛСРОК  B(1) PC(1) NA;// Кол-во
 сроков набл.
```

3) Part of the record TPOCHV description

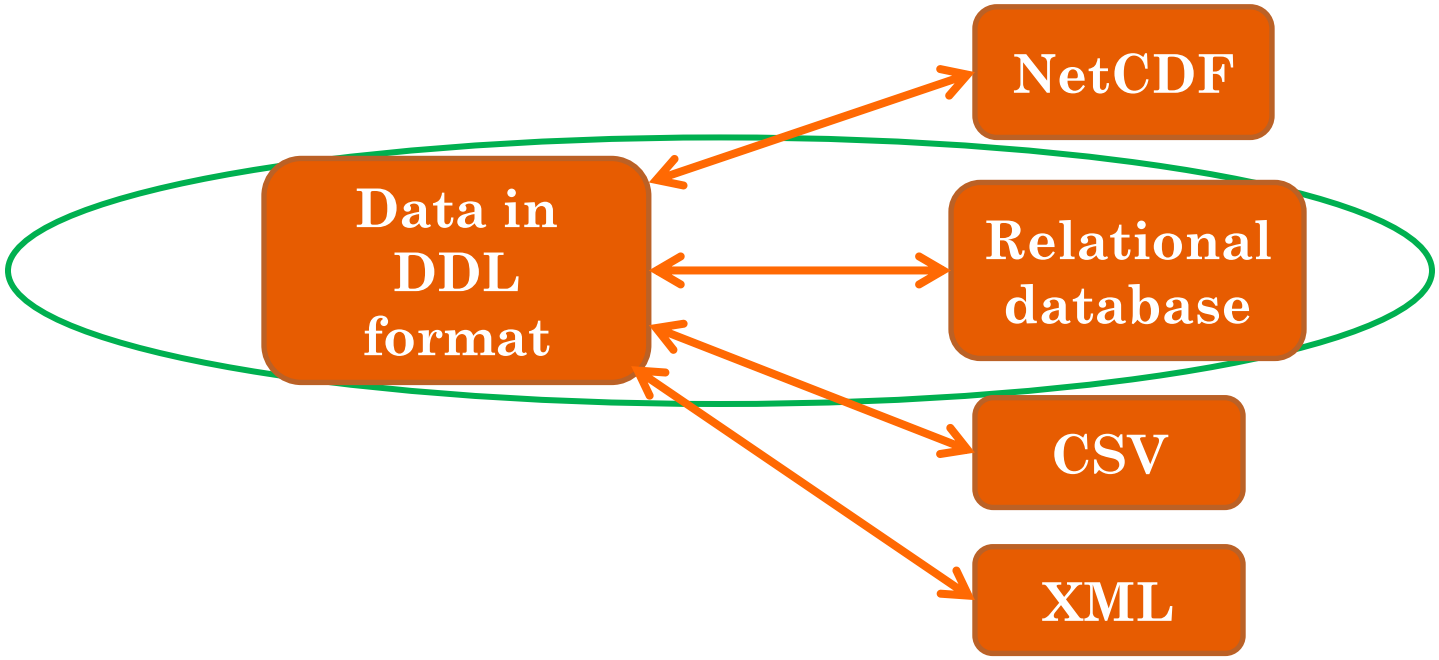
```
RBODY(3) TPOCHV ;  //
  KEY(I) ДЕНЬ     B(1) PC(2);
  CNT  СЧГРОГП  B(1) PC(1);  //
  CNT  СЧГРЕСП1 B(1) PC(1);  //
  CNT  СЧГРЕСП2 B(1) PC(1);  //
  MIT  СНЕПВЫСТ B(2) PC(4);  //
  CHA(СНЕПВЫСТ) Q  B(1) PC(1) NA;

  GRV(СЧГРОГП ) ТЕМПОГ;
    IND(1) ПРНАЛИЧ  PC(1);
     GRP SROKG;   // -- Вложенная группа
       IND(4) ГЛУБИНЫ  PC(1) ;
       MIT ТЕМПОГСТ B(2) PC(5,1) D(1); //
       CHA(ТЕМПОГСТ) Q B(1) PC(1) NA;
     END SROKG ;
   END ТЕМПОГ;
END TPOCHV;
```

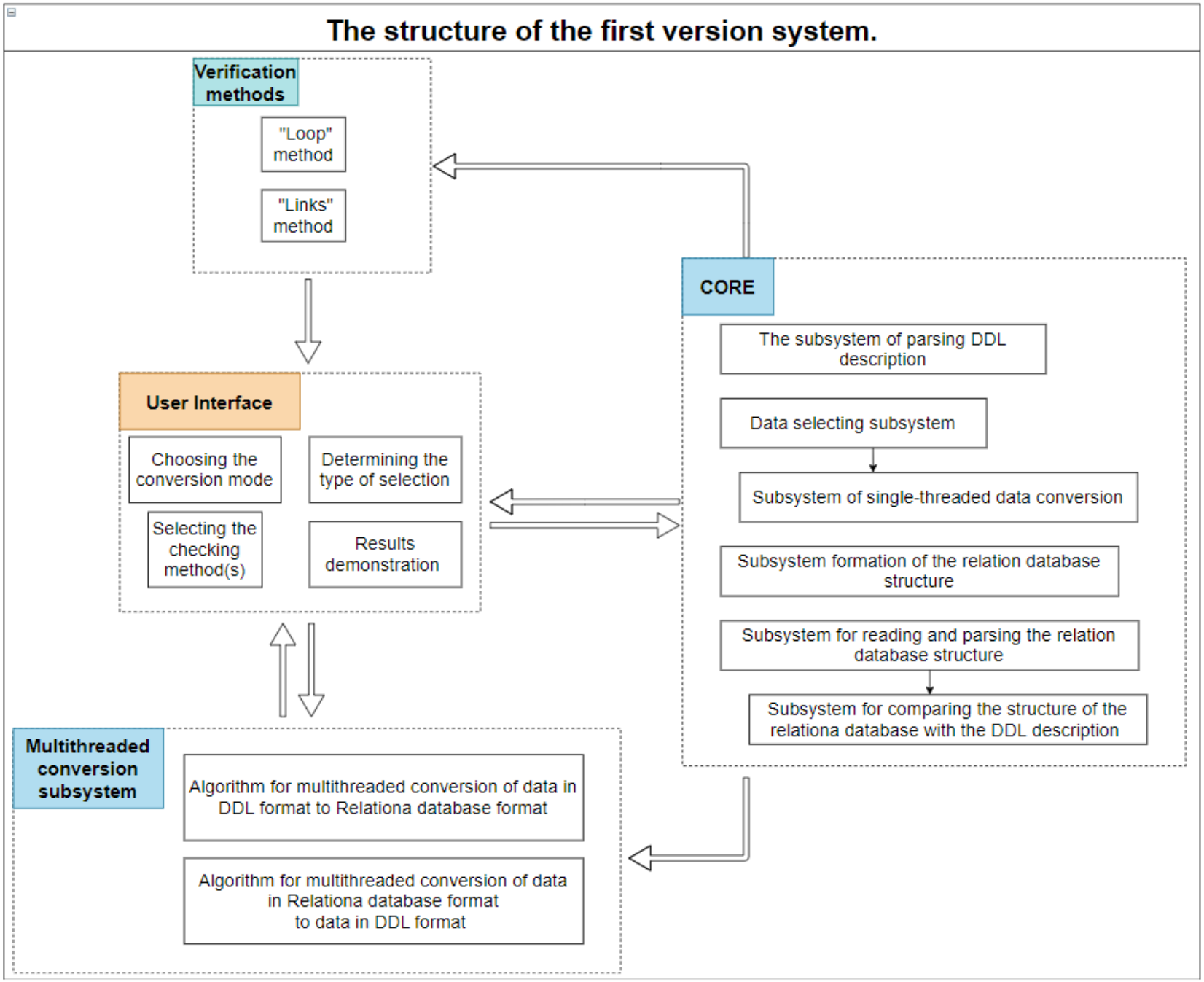
SYSTEM FOR RECIPROCAL DATA CONVERSION

The DDL format is convenient for accumulating and storing large arrays of data that make up the USDF, but using it as a data format provided to consumers is impractical due to its specificity, departmental use and complexity for use by consumers.

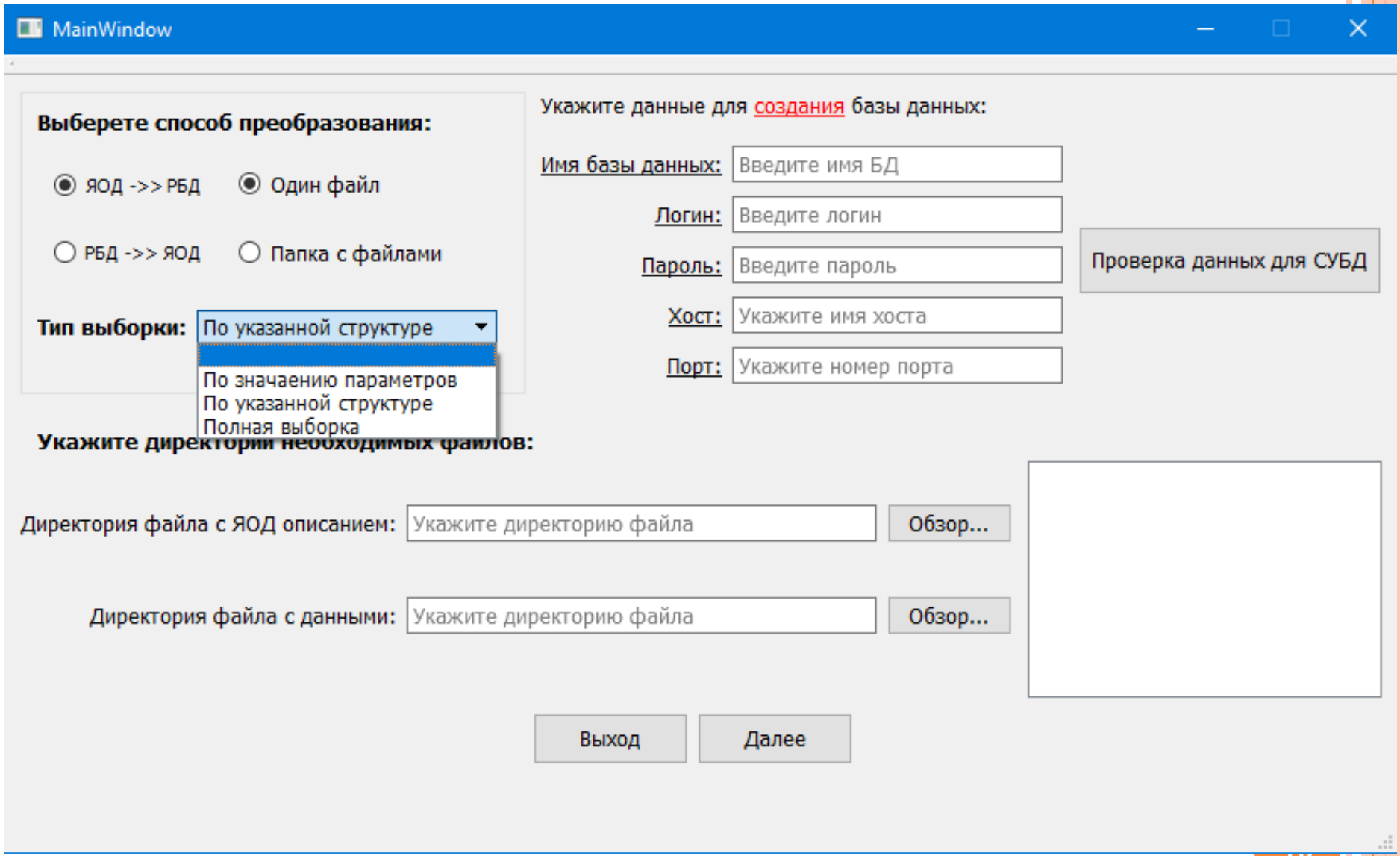
Studies have shown that to provide consumers with their information service with USDF data, the most popular formats are netCDF, XML, CSV and relational database formats (**RDB**).



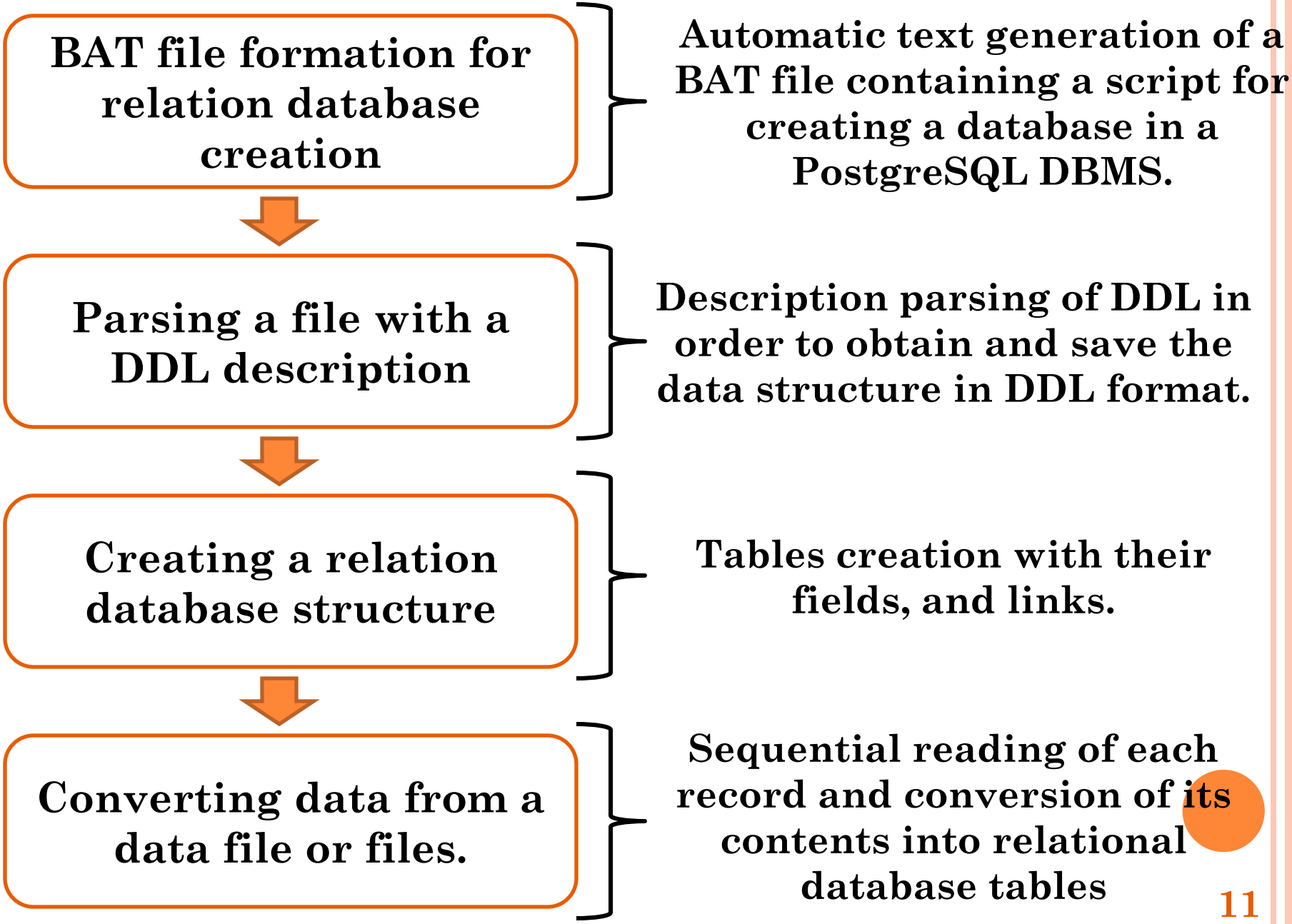
Structure of the system first version



Program interface



DDL -> RDB CONVERSION ALGORITHM



DDL -> RDB CONVERSION STAGES

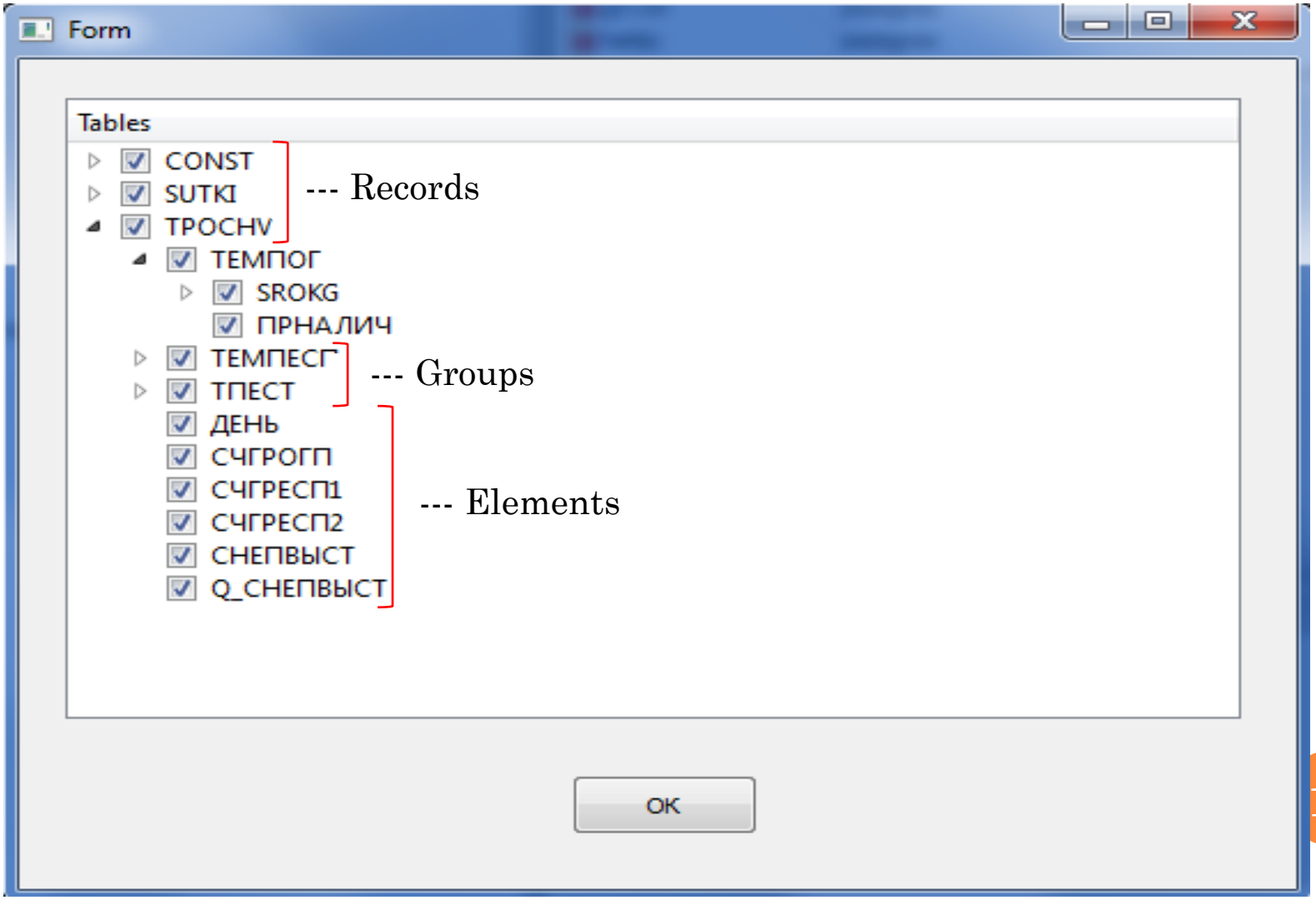
1) Automatic text generation of a BAT file containing a script for creating a database in a PostgreSQL DBMS.

BatUniCreate.bat

```
1  SET PGHOST=localhost
2  SET PGPORT=5432
3  SET PGUSER=postgres
4  SET PGPASSWORD=123
5  C:
6  cd C:\PostgreSQL\9.4\bin\\
7  createdb -O postgres "TMSS_TEST"
8  exit
```

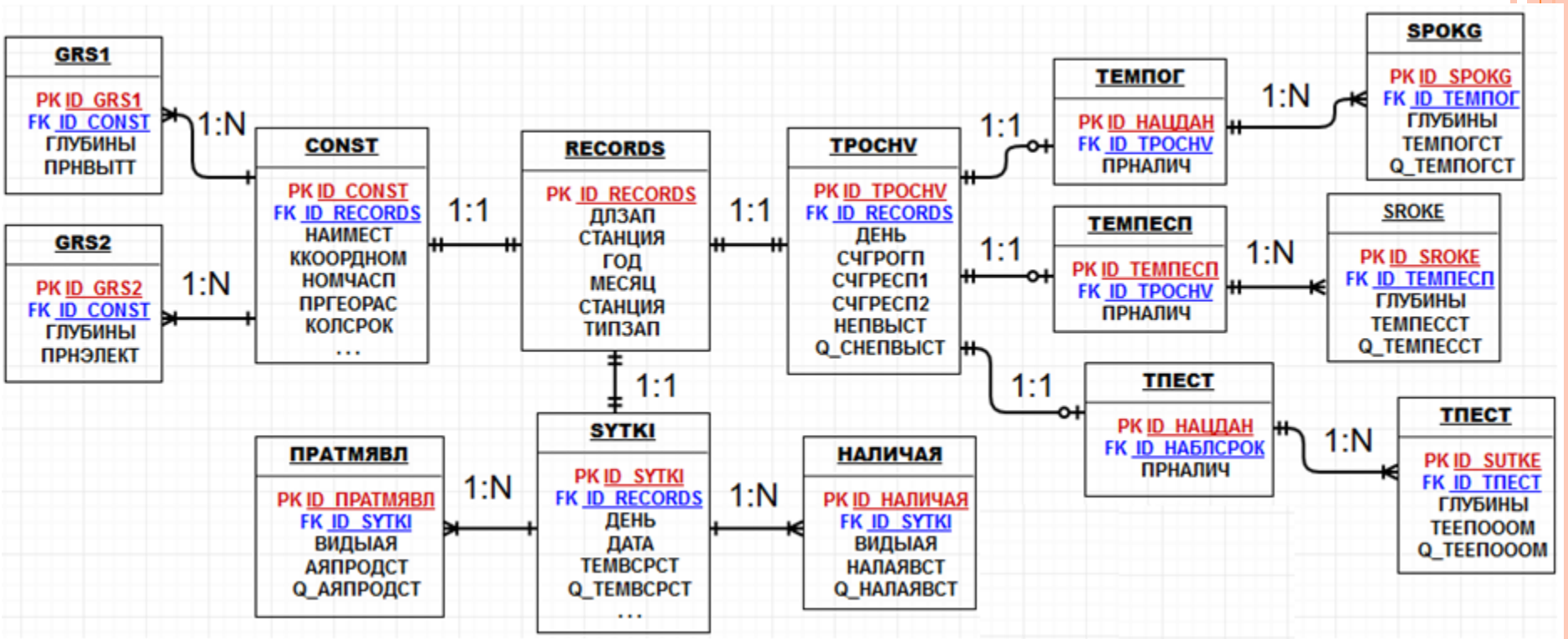
DDL -> RDB CONVERSION STAGES

2) Description parsing of DDL in order to obtain and save the data structure in DDL format.



DDL -> RDB CONVERSION STAGES

3) Relational database structure generation – the creation of tables with their fields, and relationships between tables based on the results of parsing the DDL.

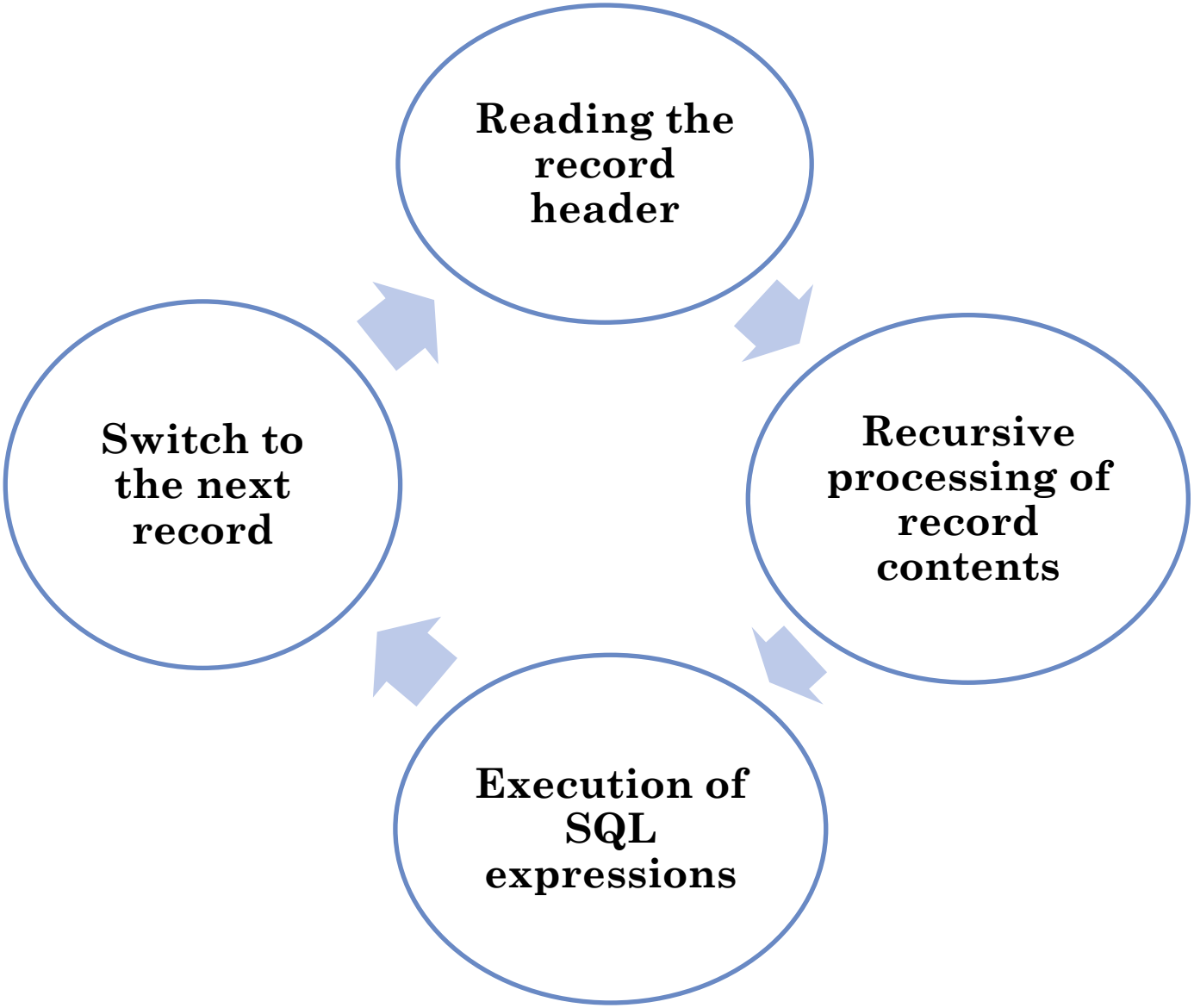


PK - Primary Key

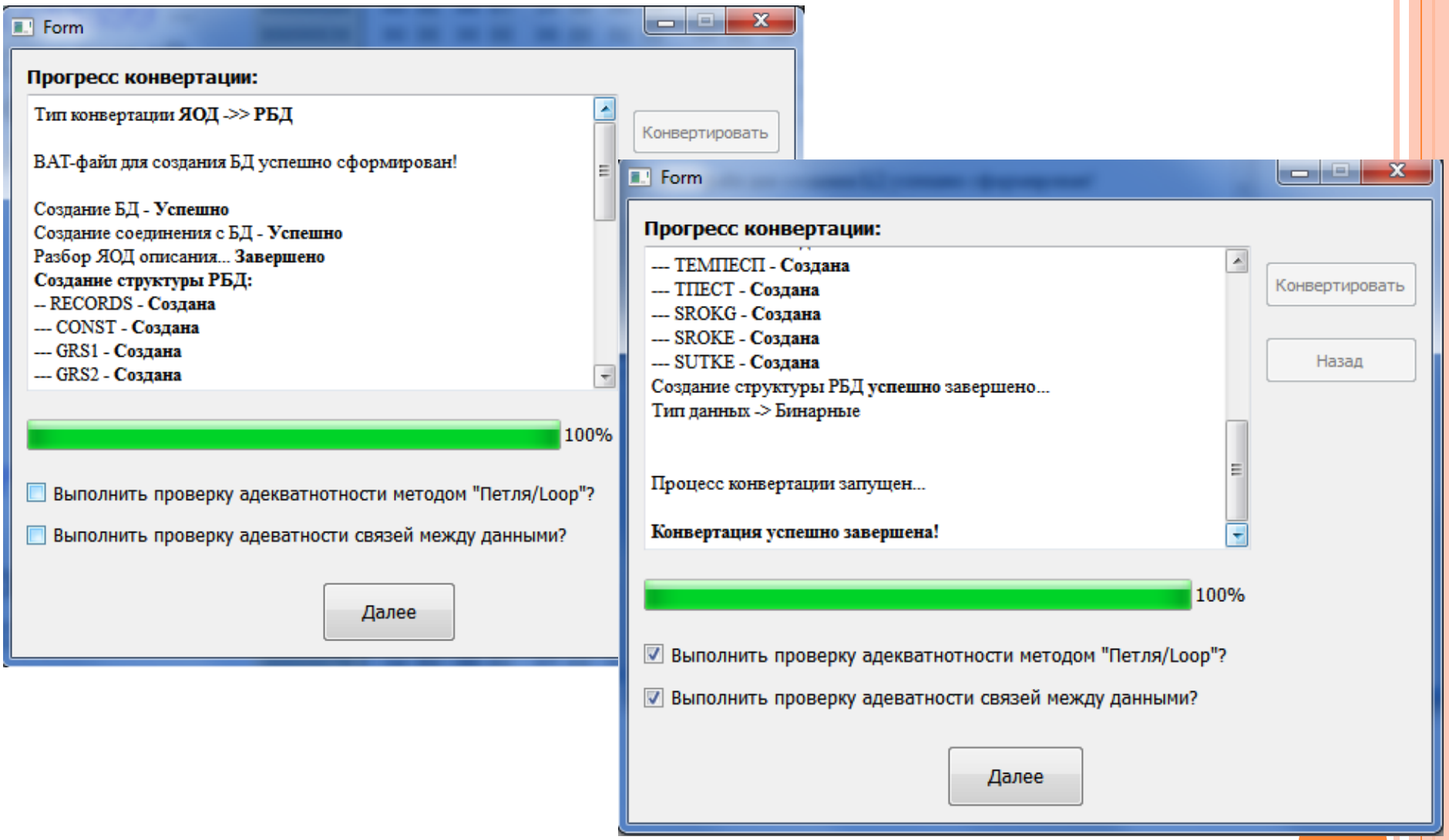
FK - Foreign Key

DDL -> RDB CONVERSION STAGES

4) Converting data from a data file or files.



Program interface



CONVERTATION RESULTS

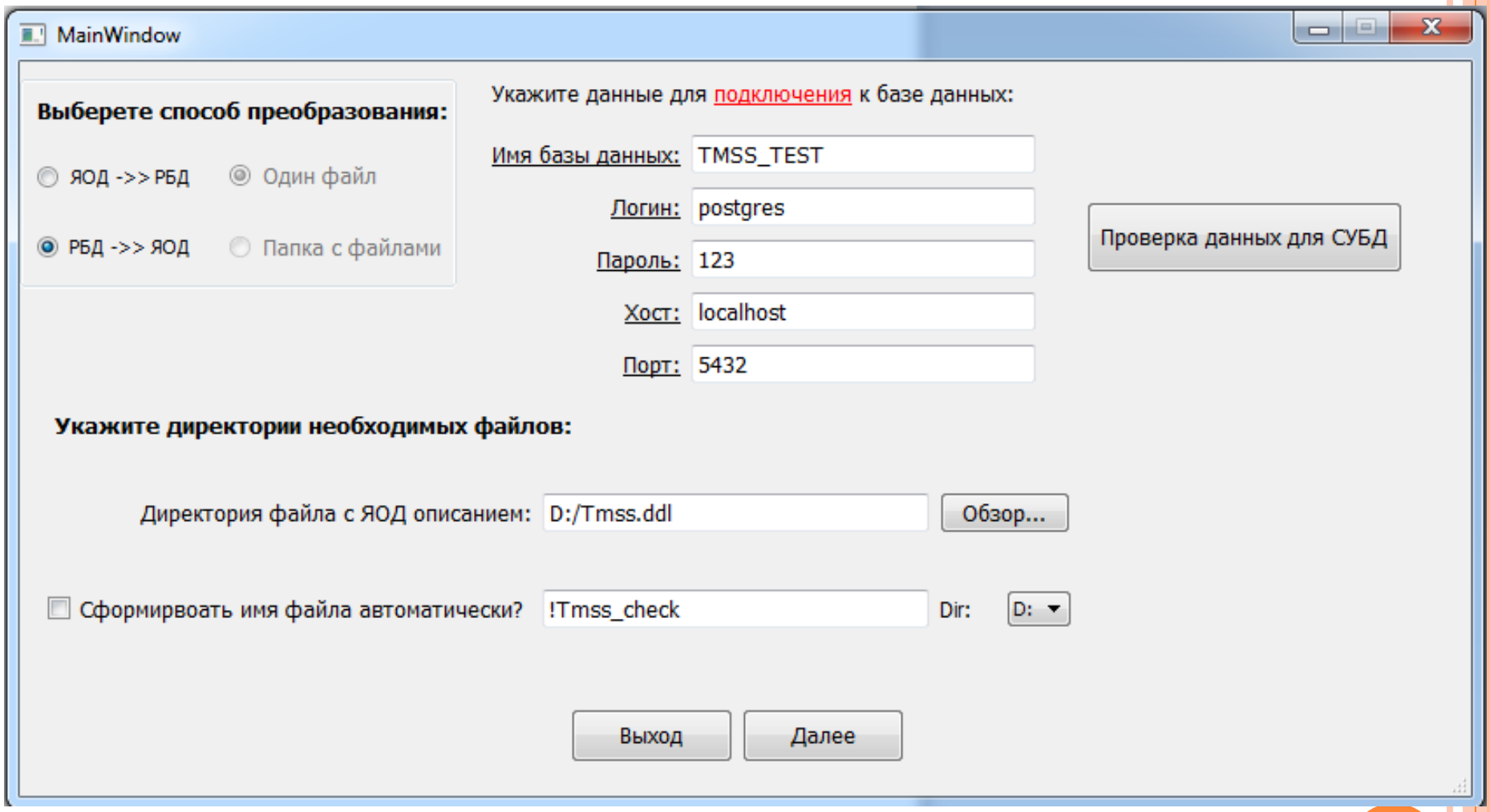
- Example of data from **RECORDS** table

	id_records [PK] serial	ДЛЗАП smallint	НУЛИ smallint	ГОД smallint	МЕСЯЦ smallint	СТАНЦИЯ integer	ТИПЗАП smallint
1	1	137	0	1985	1	7333991	1
2	2	120	0	1985	1	7333991	2
3	3	120	0	1985	1	7333991	2
4	4	120	0	1985	1	7333991	2
5	5	120	0	1985	1	7333991	2

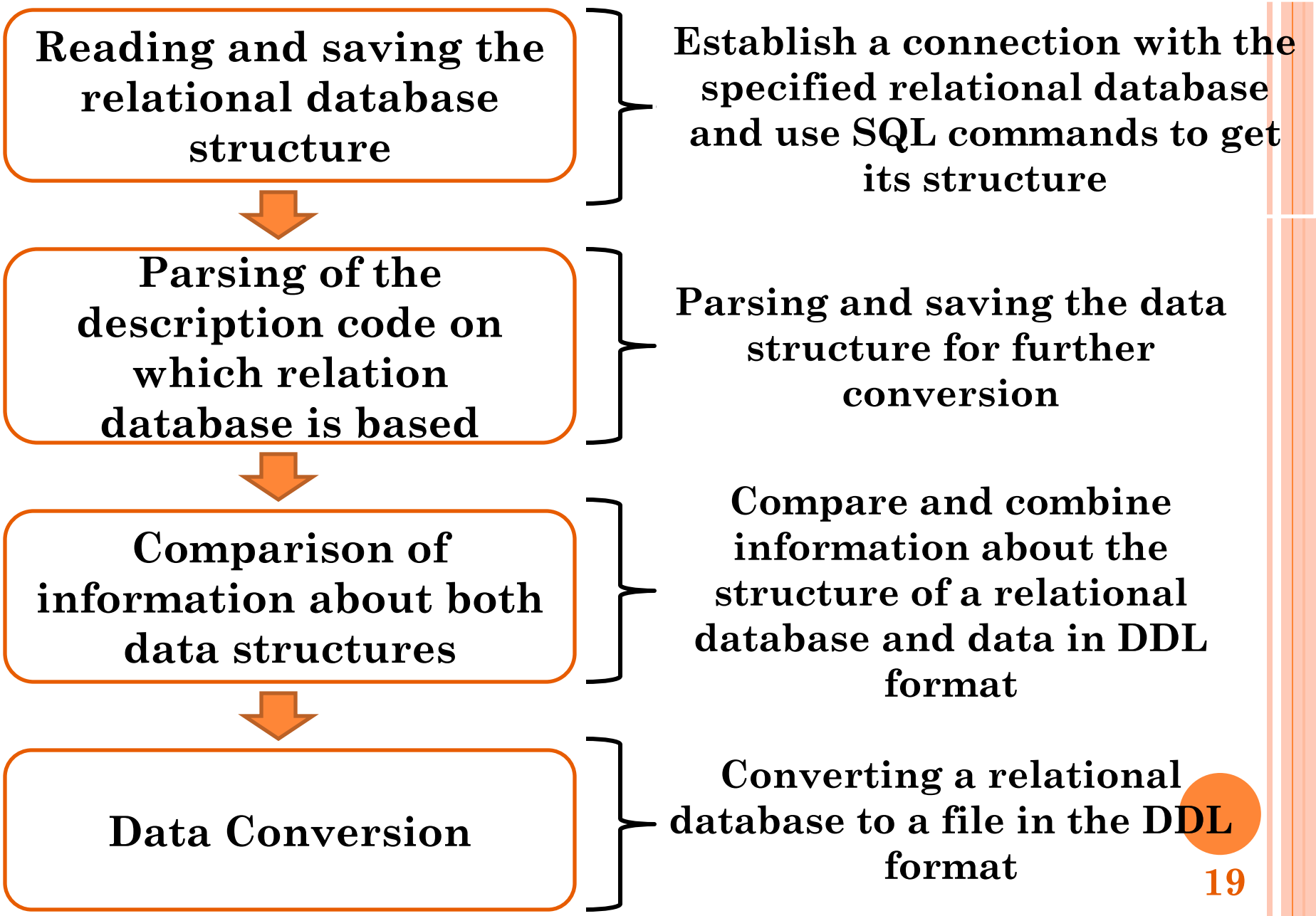
- Example of data from **SYTKI** table

	id_sutki [PK] serial	id_records bigint	ДЕНЬ smallint	ДАТА smallint	ТЕМВСРСТ numeric(5,1)	q_ТЕМВСРСТ smallint	ТЕВМАКСТ numeric(5,1)	q_ТЕВМАКСТ smallint	ТЕВМИНСТ numeric(5,1)	q_ТЕВМИНСТ smallint
1	1	2	1	1	-23.4	0	-20.1	0	-26.7	0
2	2	3	2	2	-20.7	0	-19.6	0	-23.4	0
3	3	4	3	3	-24.0	0	-20.7	0	-26.0	0
4	4	5	4	4	-26.7	0	-25.2	0	-27.4	0
5	5	6	5	5	-29.2	0	-27.3	0	-30.7	0
6	6	7	6	6	-27.5	0	-26.3	0	-30.1	0
7	7	8	7	7	-28.3	0	-26.1	0	-31.3	0

Program interface



RDB -> DDL CONVERSION ALGORITHM



METHODS OF ADEQUACY CONTROL

The adequacy control subsystem includes the following methods:

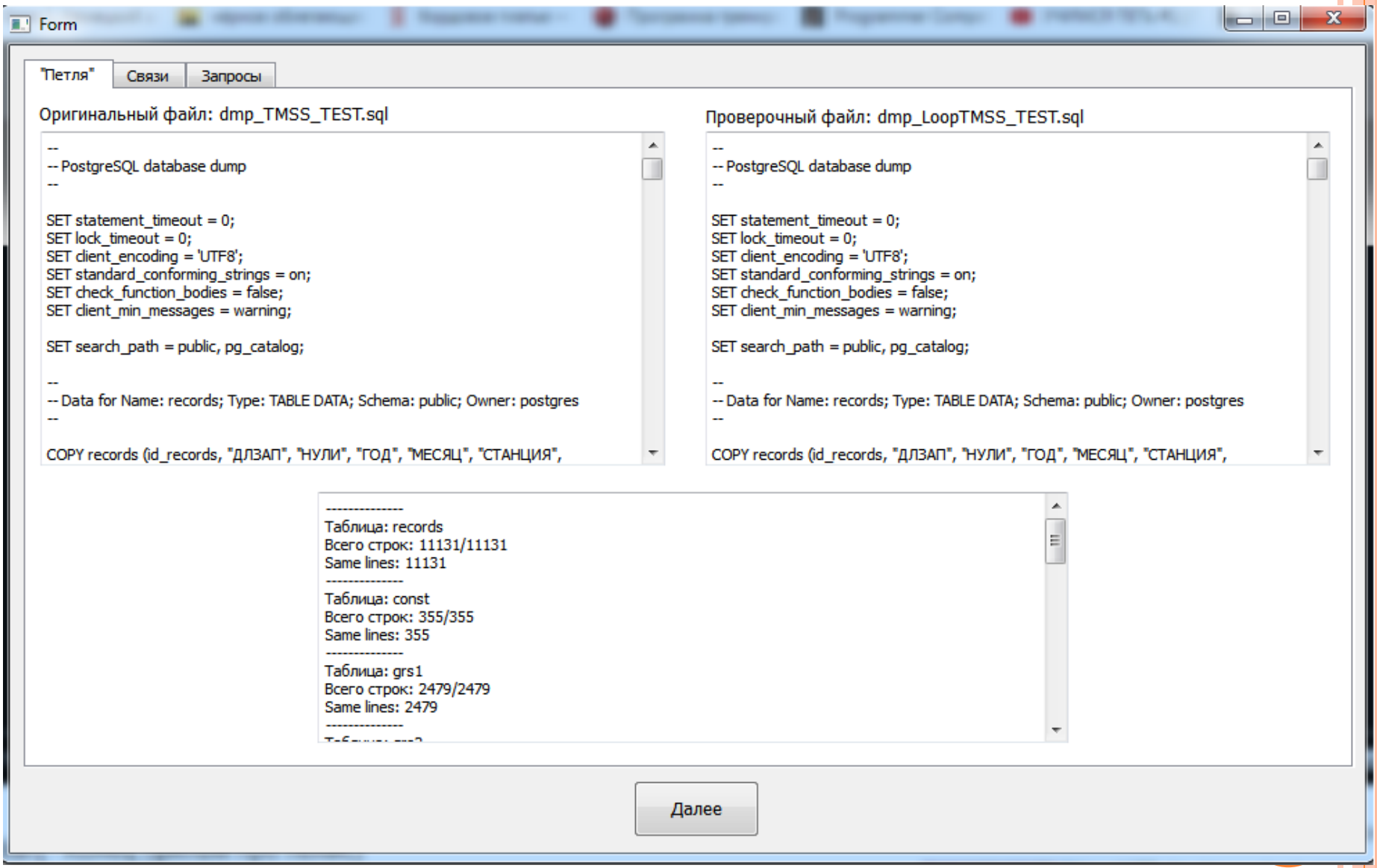
1) **"Loop" – after the conversion is completed, the reverse conversion is performed, followed by a comparison of the results;**
2) **Comparison of the results of adequate data queries;**
3) **Comparison of relationships between data in different models;**

Conversion results

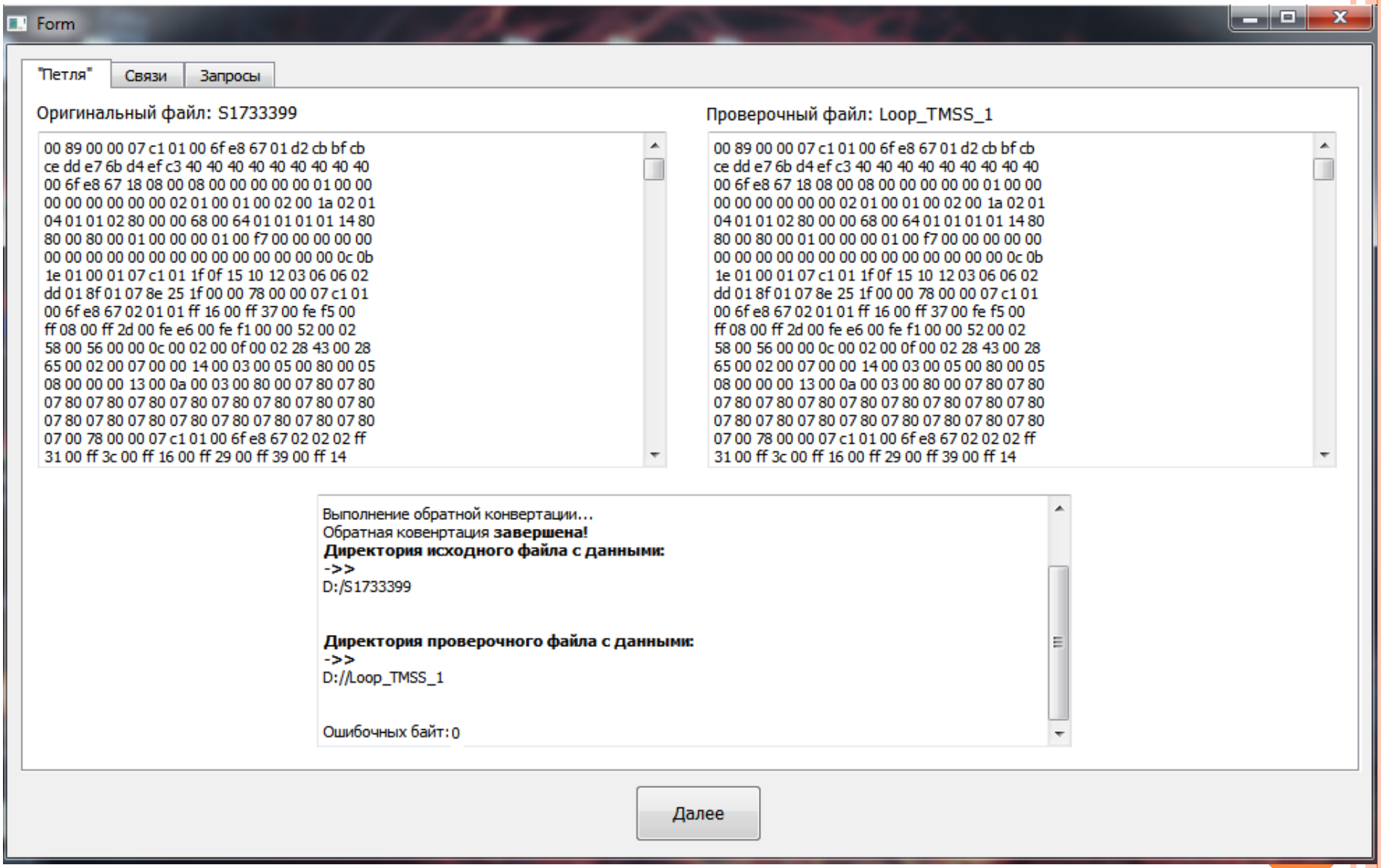
!test_tmss

00000000	00	01	02	03	04	05	06	07	08	09	0a	0b	0c	0d	0e	0f	
00000000	00	89	00	00	07	c1	01	00	6f	e8	67	01	d2	cb	bf	cb	.‰...Б..оиg.ТЛїЛ
00000010	ce	dd	e7	6b	d4	ef	c3	20	20	20	20	20	20	20	20	20	ОЭзkФпГ
00000020	00	6f	e8	67	18	08	00	08	00	00	00	00	00	01	00	00	.оиg.....
00000030	00	00	00	00	00	00	02	01	00	01	00	02	00	1a	02	01	.....
00000040	04	01	01	02	80	00	00	68	00	64	01	01	01	01	14	80	....Ъ..h.d.....Ъ
00000050	80	00	80	00	01	00	00	00	01	00	f7	00	00	00	00	00	Ъ.Ъ.....ч.....
00000060	00	00	00	00	00	00	00	00	00	00	00	00	00	00	0c	0b	.....
00000070	1e	01	00	01	07	c1	01	1f	0f	15	10	12	03	06	06	02	.....Б.....
00000080	dd	01	8f	01	07	8e	25	1f	00	00	78	00	00	07	c1	01	Э.Ц..Ћ%...x...Б.
00000090	00	6f	e8	67	02	01	01	ff	16	00	ff	37	00	fe	f5	00	.оиg...я..я7.юх.
000000a0	ff	08	00	ff	2d	00	fe	e6	00	fe	f1	00	00	52	00	02	я..я-.юж.юс..R..
000000b0	58	00	56	00	00	0c	00	02	00	0f	00	02	28	43	00	28	X.V.....(C.(
000000c0	65	00	02	00	07	00	00	14	00	03	00	05	00	80	00	05	e.....Ъ..
000000d0	08	00	00	00	13	00	0a	00	03	00	80	00	07	80	07	80	.....Ъ..Ъ.Ъ
000000e0	07	80	07	80	07	80	07	80	07	80	07	80	07	80	07	80	.Ъ.Ъ.Ъ.Ъ.Ъ.Ъ.Ъ.Ъ
000000f0	07	80	07	80	07	80	07	80	07	80	07	80	07	80	07	80	.Ъ.Ъ.Ъ.Ъ.Ъ.Ъ.Ъ.Ъ
00000100	07	00	78	00	00	07	c1	01	00	6f	e8	67	02	02	02	ff	..x...Б..оиg...я
00000110	31	00	ff	3c	00	ff	16	00	ff	29	00	ff	39	00	ff	14	1.я<.я..я).я9.я.
00000120	00	ff	18	00	00	69	00	02	58	00	51	00	00	0e	00	02	.я...i..X.Q.....
00000130	00	18	00	02	28	54	00	28	76	00	02	00	02	00	00	31	....(T.(v......1
00000140	00	07	00	0b	00	80	00	05	08	00	00	00	13	00	0a	00	.....Ъ.....
00000150	03	00	80	00	07	80	07	80	07	80	07	80	07	80	07	80	..Ъ..Ъ.Ъ.Ъ.Ъ.Ъ.Ъ
00000160	07	80	07	80	07	80	07	05	00	80	07	80	07	80	07	80	.Ъ.Ъ.Ъ...Ъ.Ъ.Ъ.Ъ
00000170	07	80	07	80	07	80	07	80	07	00	78	00	00	07	c1	01	.Ъ.Ъ.Ъ.Ъ..x...Б.
00000180	00	6f	e8	67	02	03	03	ff	10	00	ff	31	00	fe	fc	00	.оиg...я..я1.юь.
00000190	fe	fc	00	ff	2a	00	fe	de	00	fe	eb	00	00	4c	00	02	юь.я*.юЮ.юл..L..

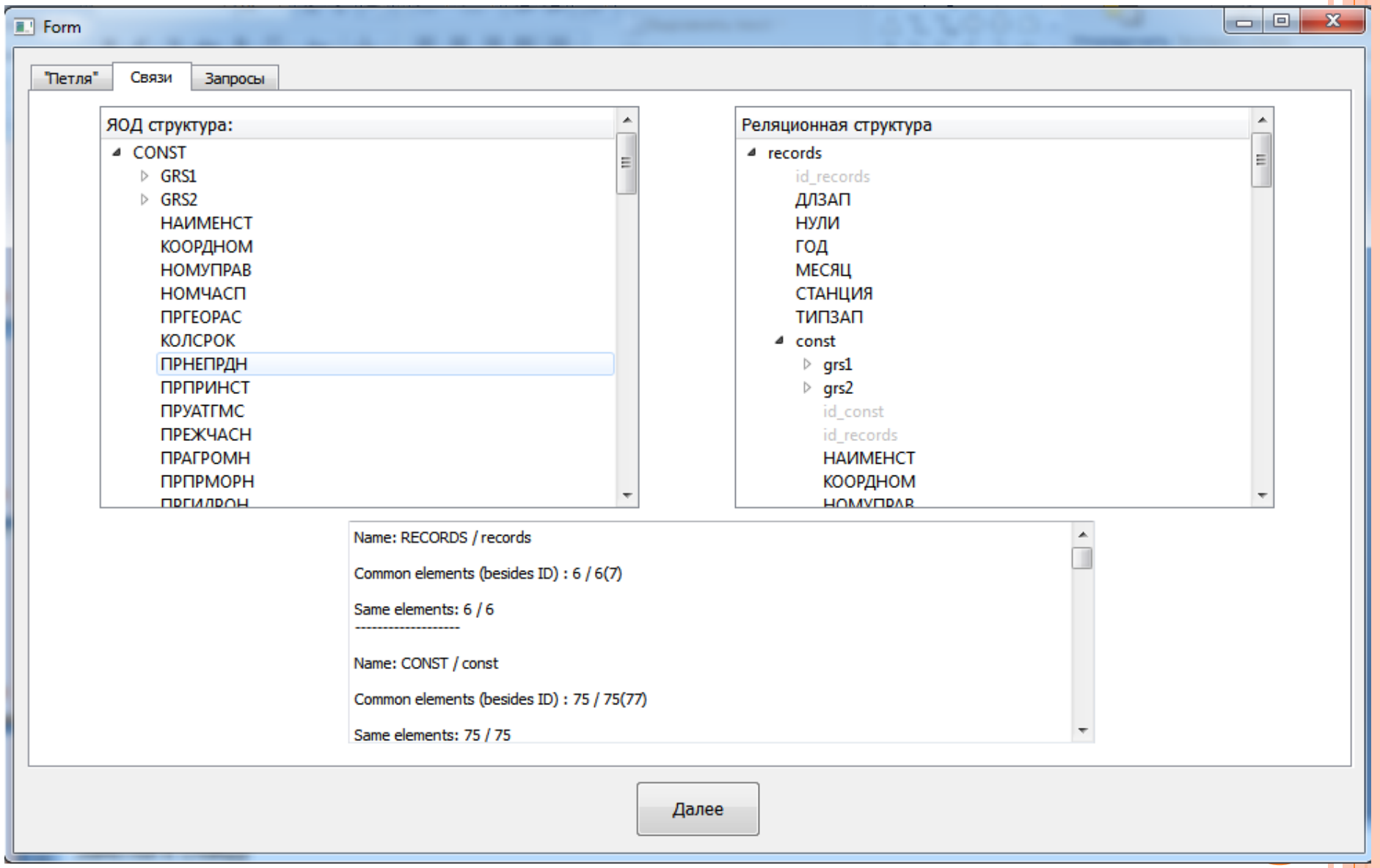
Adequacy of results



Adequacy of results



Adequacy of results



Conclusion

- **Major results:**
  - **The system of mutual data conversion has been designed, with a subsystem monitoring the adequacy of the conversion performed.**
  - **The first version of the mutual conversion system has been developed and implemented programmatically.**
  - **Methods of adequacy verification have been developed.**
  - **The first version of the system was tested using the example of aerological and meteorological data.**

Thank you for attention!