

HISTORICAL PHENOLOGICAL DATA MIGRATION

Helena Lebo Andreis
 Meteorological and Hydrological Service of Croatia (DHMZ)



ABOUT THE PHENOLOGICAL DATABASE

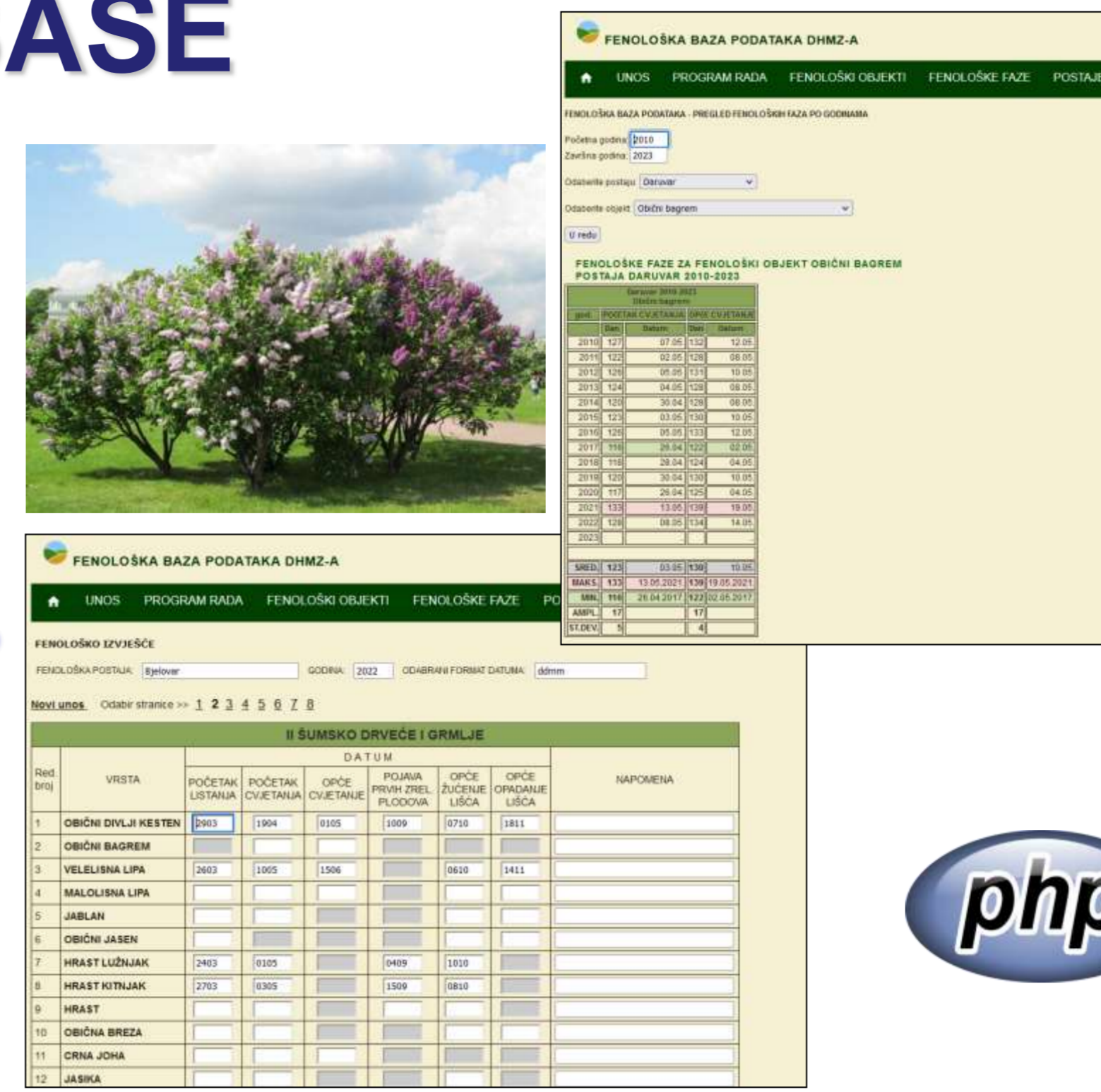
Several years ago, new phenological relational database was developed at DHMZ.

Type of database is **PostgreSQL**. All applications were developed in **PHP, JavaScript and HTML**.

The database users access the database through a browser.

In the first stage, the data from **1995 to present** were stored to the database.

At that time, the data from **1951 – 1994** remained stored in the historical file system, which had to be migrated to the relational database.



DATA MIGRATION

The phenological data between **1951-1994** were migrated from the historical file system to the relational database. Data migration was a complex and time-consuming task due to several reasons, mostly due to incompatible IDs and data format.

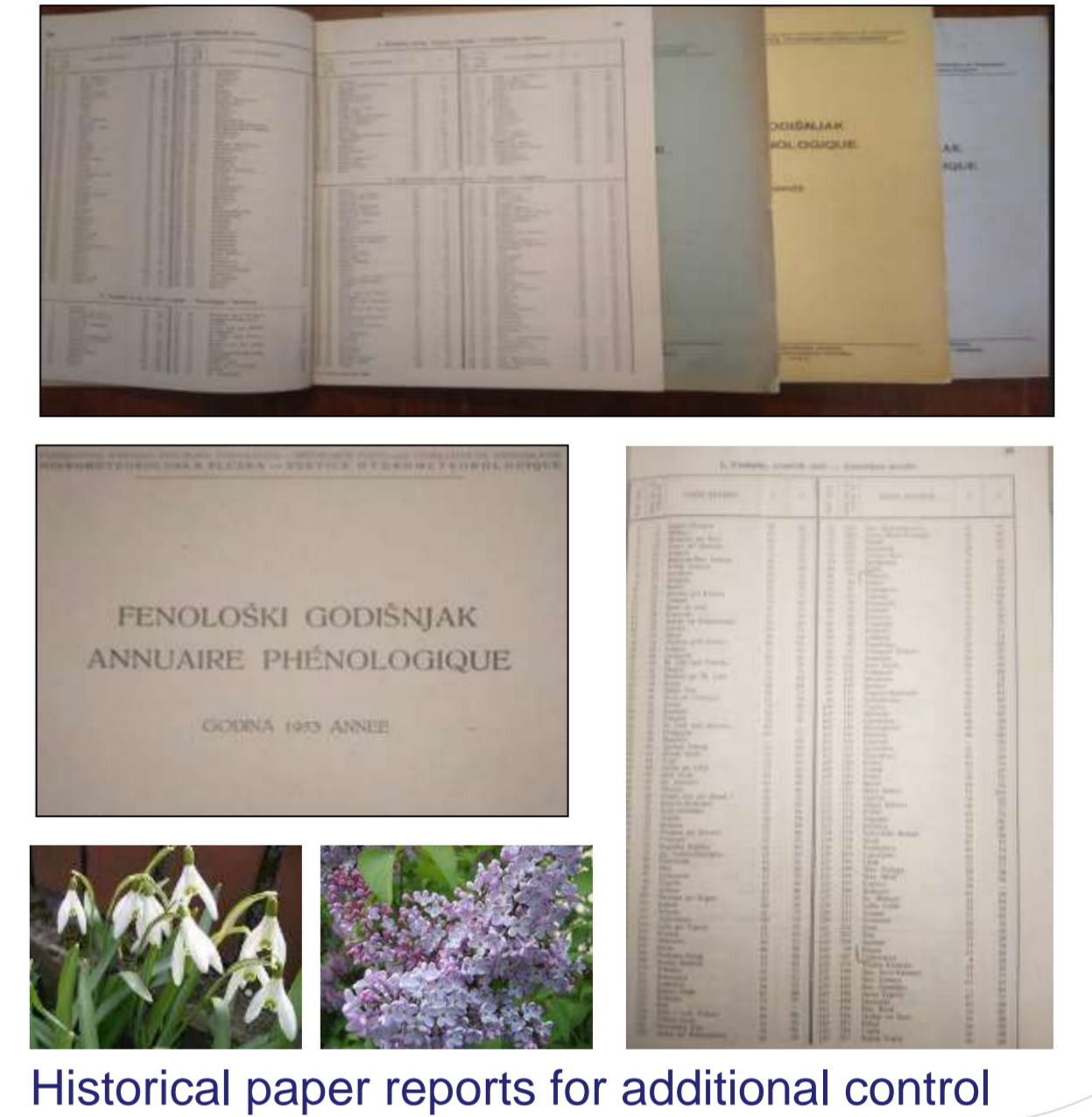
CHALLENGES

Problem	Solution
Incompatible station IDs	Finding matching pairs of station IDs between database and file system
Incompatible plant IDs	Finding matching pairs of plant IDs between database and file system
Incompatible phenological stage IDs	Finding matching pairs of stage IDs between database and file system
Incompatible data format	Creating compatible CSV files for importing to database
Many historical plants that are no longer observed	Inserting missing plants to the database catalogue

METHODS

Methods or tools
PHP programming language
Bash
SQL
Manual corrections
Comparison with historical paper reports
Lots of team work and communication

ADDITIONAL CONTROL



Historical paper reports for additional control

RESULTS

The result of phenological data migration is a completed dataset from **1951-2023**.

Data	Number of stations	Number of plants	Number of phenological stages	Number of records	Dataset
Migrated data from old file system	123	370 (271 historical)	42	Around 400 000	1951-1994
Total data in database after migration	152	964	45	Around 670 000	1951-2023

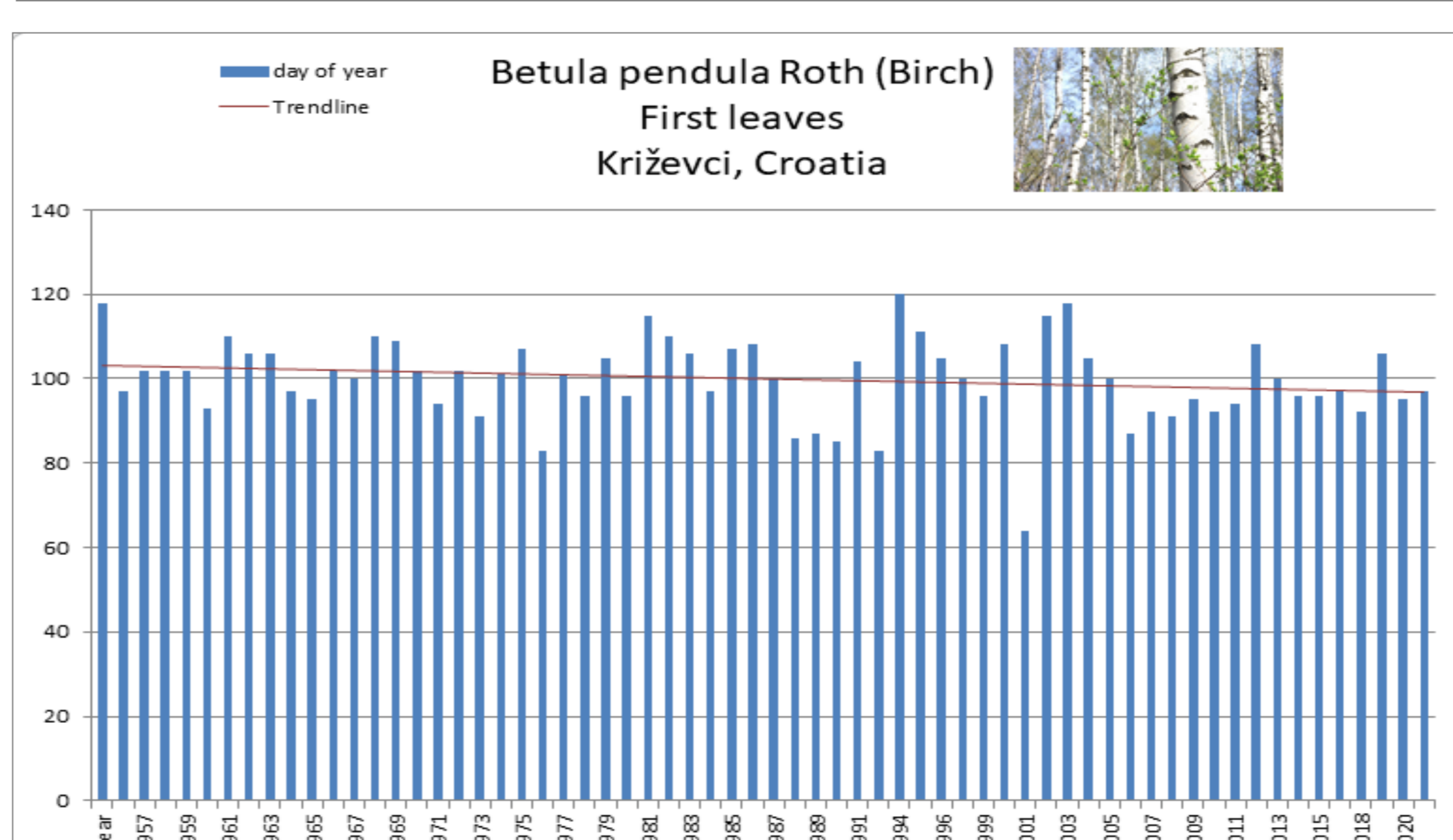
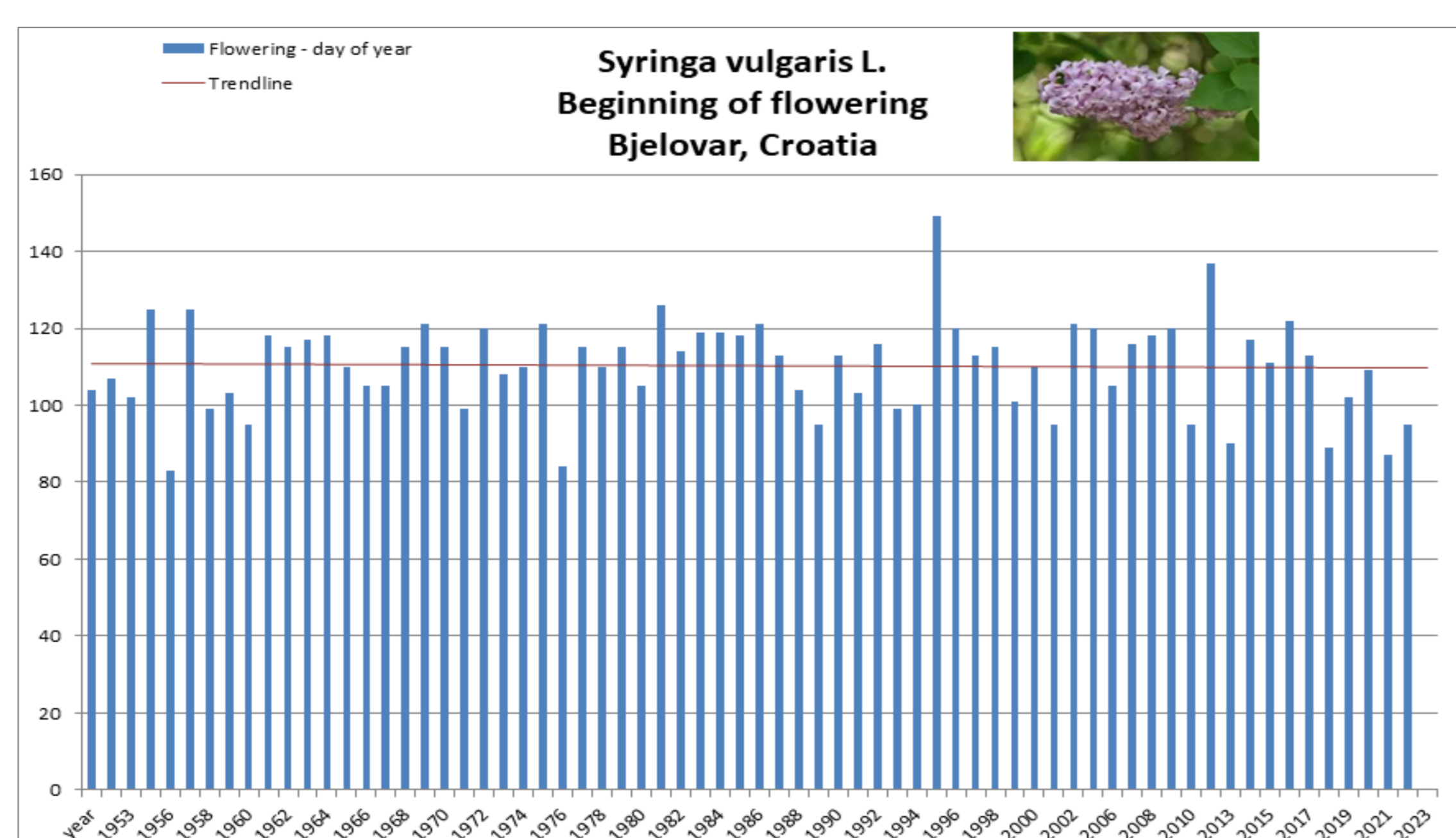
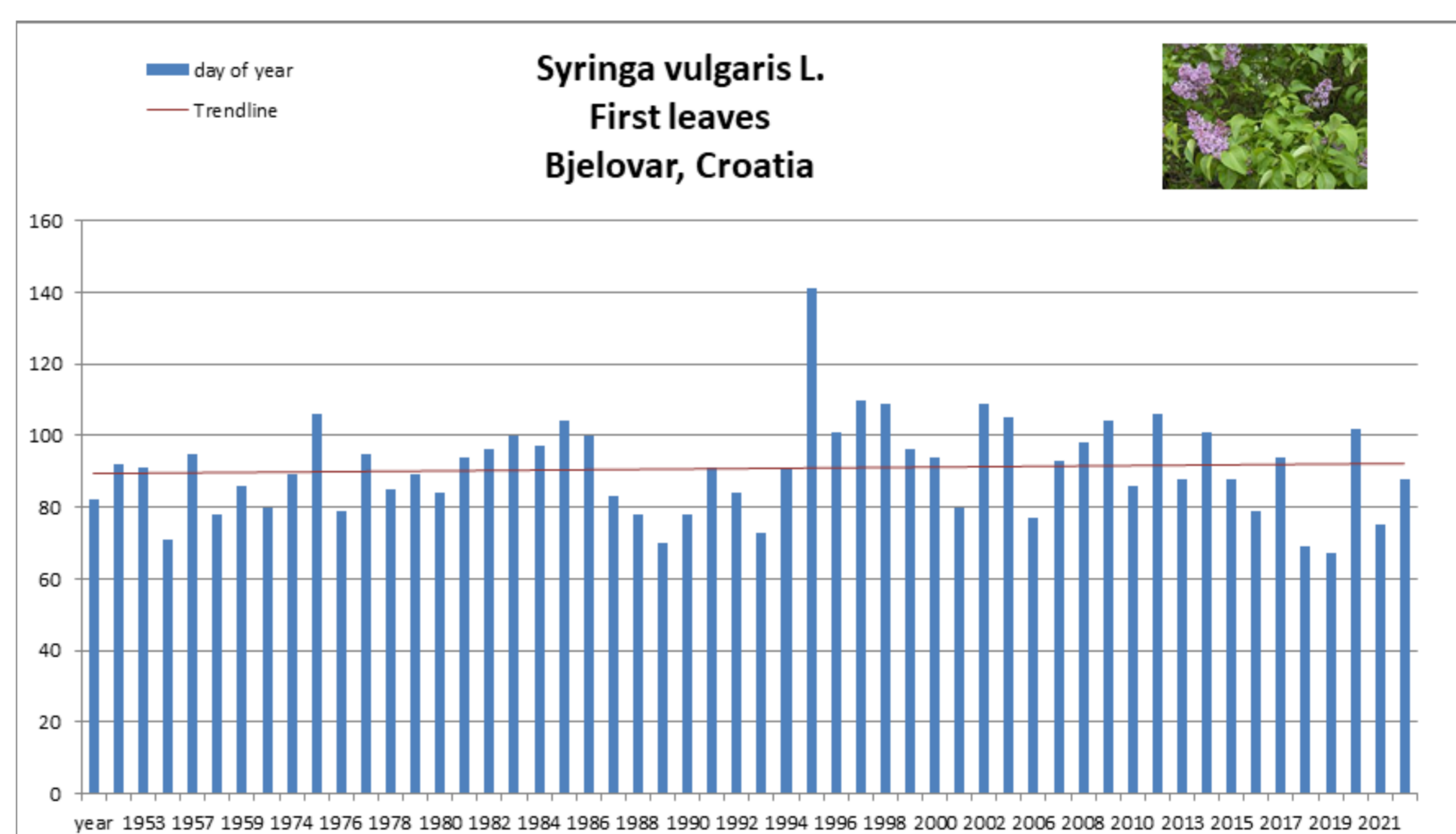
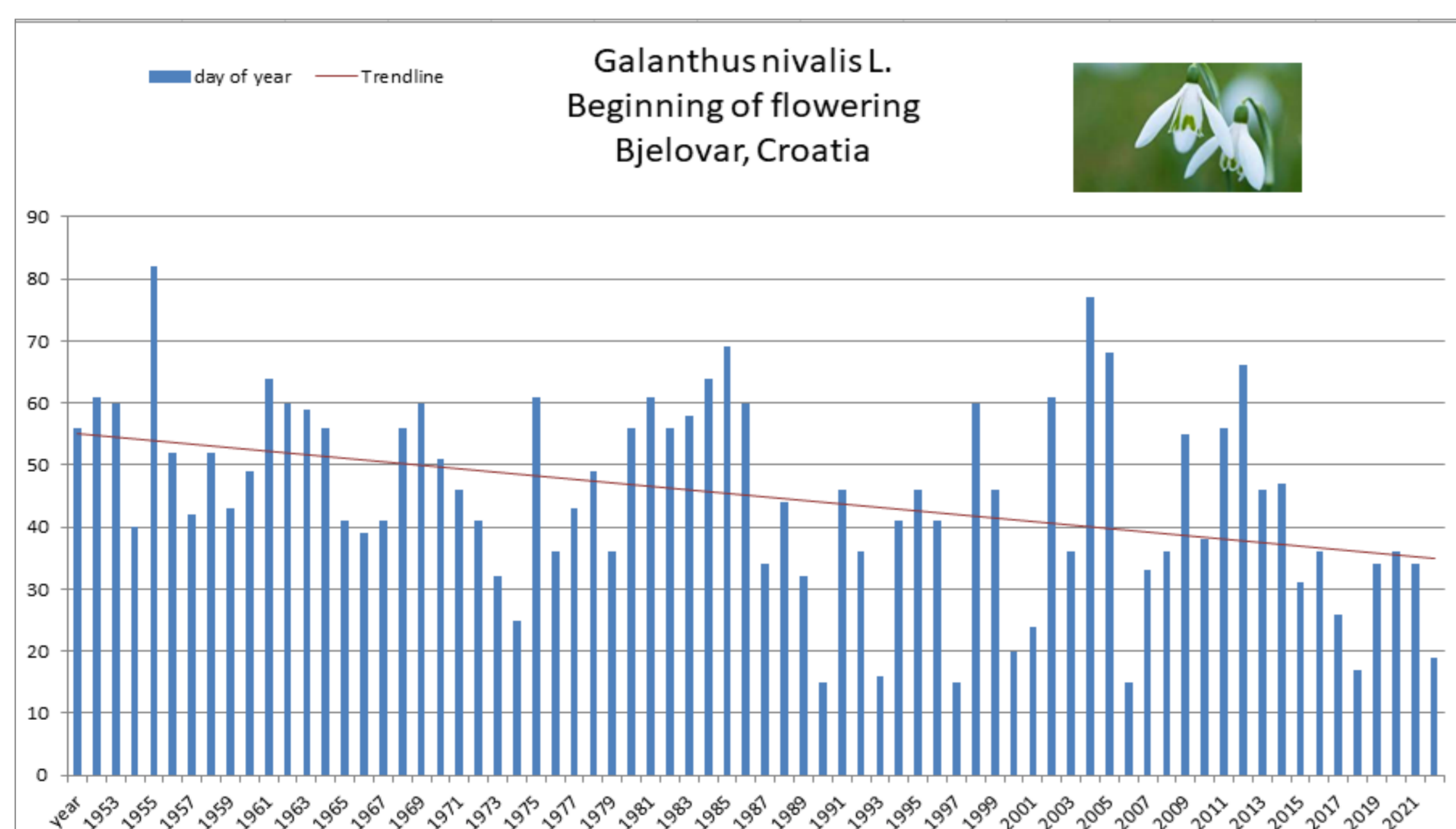
THE OLDEST RECORD?

11.01.1951.



The beginning of flowering of Common hazel (*Corylus avellana* L.) at Topusko

Examples of completed datasets:



TEAM

Data migration is a result of good communication and mutual work of the file system developer and the relational database developer.

