

Information-computational system for storage and search of environmental datasets based on the Semantic Web technologies

Titov A.G, Gordov E.P., Okladnikov I.G.

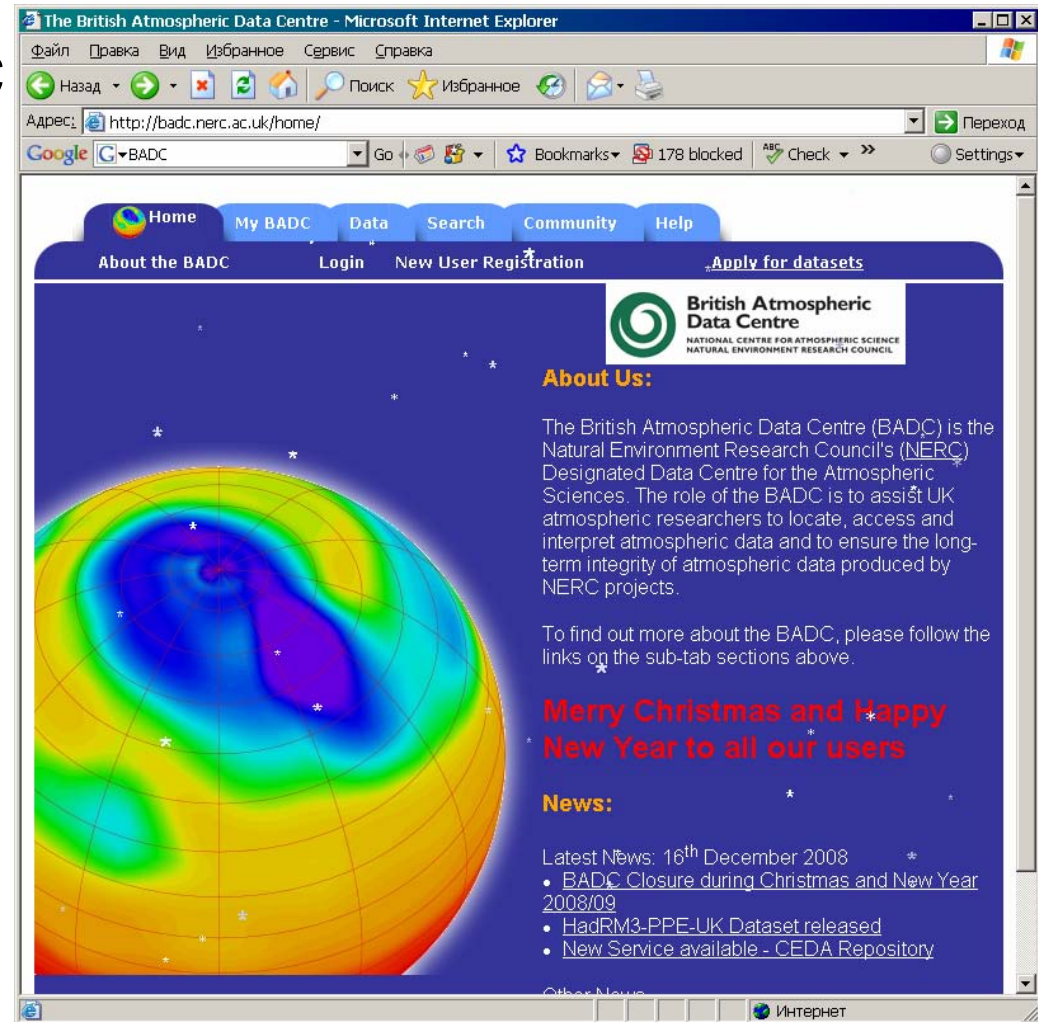
Institute of Monitoring of Climatic and Ecological
Systems SB RAS / SCERT, Tomsk, Russia

Introduction

- Earth Science: numerous archives of heterogeneous data, often huge in size
- Dataset use and re-use problem
- Task of search and operative access to relevant data and results of data processing taking into account subject domain specifics

Overview of existing systems

- British Atmospheric Data Center
 - Atmospheric data
 - Search functionality
 - Data access or application



<http://badc.nerc.ac.uk/home/index.html>

NERIN

■ Northern Eurasia Regional Information Network

- Metadata database
- Metadata input/update interface
- Search functionality

NERIN Metadata Entry Form - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Address <http://wwwdata.forestry.oregonstate.edu/MDEDIT/teamform.aspx?mode=xml&mdi=95> Go Links

Dataset Title:
my test dataset

Investigators:

Investigator: Alexander G Titov ; any ; tag27@mail.ru [Delete](#) | [Edit](#)

First Name* Middle Name : Last Name*

Affiliation*

Email* [More Investigators](#)

Metadata Authors:

Saved Author: Alexander G Titov ; tag27@mail.ru ; phone: [Delete](#) | [Edit](#)

First Name* Middle Name: Last Name*

Phone : Email* [More Authors](#)

NEESPI Science Component:

<input type="checkbox"/> Carbon Dynamics	<input checked="" type="checkbox"/> Forest ecosystems	<input type="checkbox"/> Other Biogeochemical cycles
<input type="checkbox"/> Cold land processes	<input type="checkbox"/> Grasslands and agricultural ecosystems	<input type="checkbox"/> Tundra ecosystems
<input type="checkbox"/> Coastal zone	<input type="checkbox"/> Human dimensions	<input type="checkbox"/> Wetlands and other aquatic systems
<input type="checkbox"/> Energy and water cycles	<input type="checkbox"/> Land use and Land Cover	

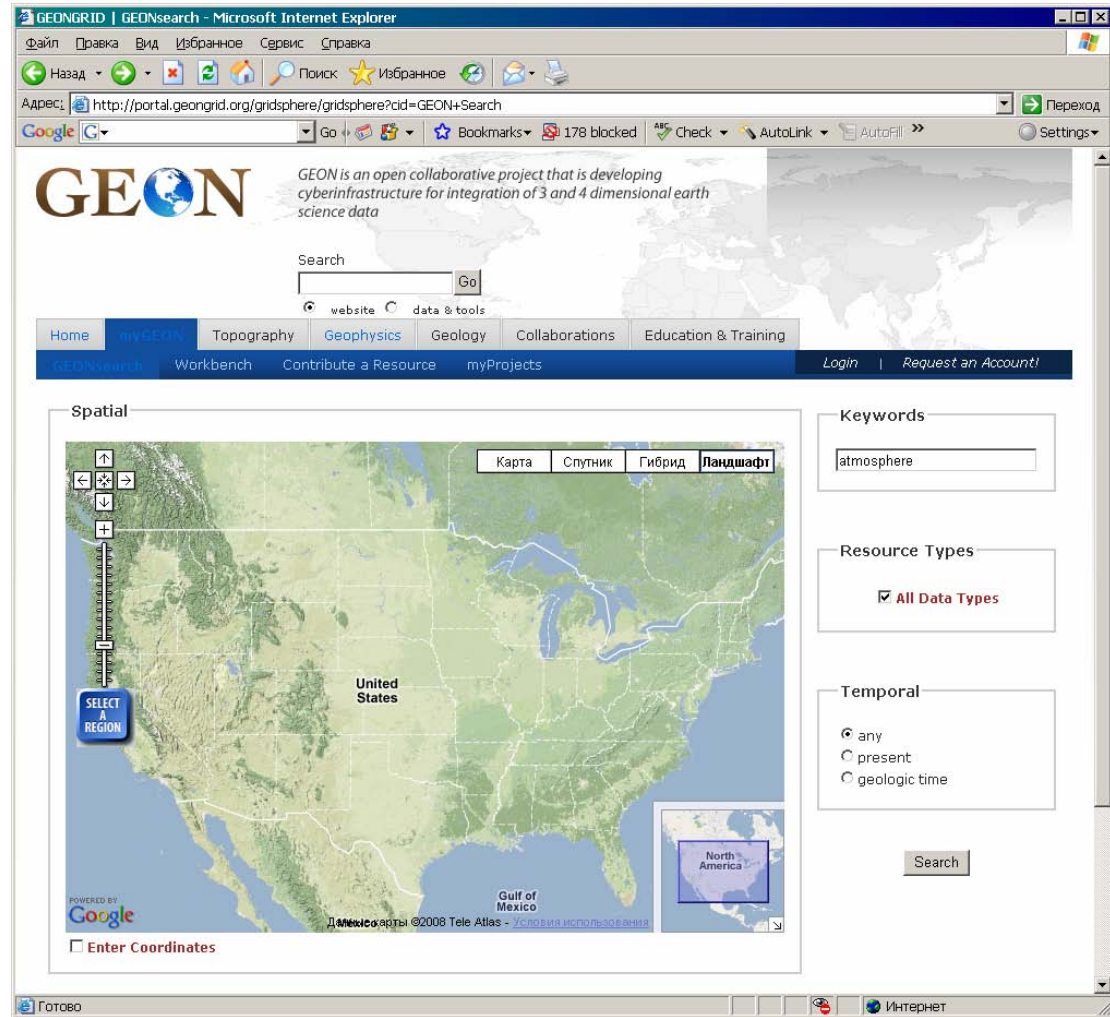
Dataset Information:

Start Date*	20041001	yyyymmdd	End Date* :	20041004	yyyymmdd
Last Modified:	20041004	yyyymmdd	Temporal Resolution:	Monthly	
Spatial Resolution:					

<http://nerin.scert.ru/>

GEON

"...a future in which someone can sit at a terminal and have easy access to vast stores of data of almost any kind, with the easy ability to visualize, analyze and model those data."



<http://www.geongrid.org/>

Home
PN Administration

Information

Projects in ISDC
Product Types
Documents
Software
FAQ
Links
Education
Statistics

Data Access

How to get data

MyRetrieval
Product Browser
MyDirectory
MyRequests

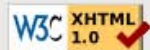
MyAdmin
CHAMP-Admin
MyMonitor

Collaboration

MyAccount
Forum
Contact

There are 1 registered user online.

You are logged-in as rit.



Report Bug

Welcome to the Information System and Data Center for geoscientific data



ISDC's online service portal is an access point for all manner of geoscientific geodata, its corresponding metadata, scientific documentation and software tools. The majority of the data and information, the portal currently offers to the public, are global geomonitoring products such as satellite orbit and Earth gravity field data as well as geomagnetic and atmospheric data for the exploration. These products for Earth's changing system are provided via state-of-the-art retrieval techniques. The ISDC portal's design and the operation is a project of the ISDC-team within the GFZ's Data Center. We invite you to use our services and they will benefit your scientific work. If you have any queries while using the site, please use the -Symbol to get context-sensitive help. Try it now .

Personal Block

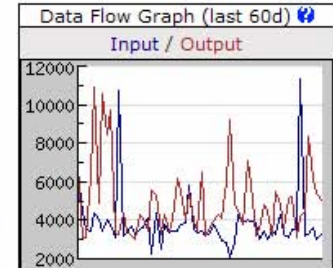
favorite product types :

- CH-AI-3-ATM
- CH-AI-3-IVP
- CH-AI-3-TCR
- CH-AI-3-WVP

Request Limits (24h)

count (50/2000)

size (102 MB/2 GB)



- Forum
- ▶ **ISDC**
 - ▶ General discussions
 - ▶ Bug-Reports
 - ▶ Questions
 - ▶ **CHAMP**
 - ▶ Atmosphere/Ionosphere
 - ▶ Magnetic/Electric Field
 - ▶ Orbit/Gravity
 - ▶ General
 - ▶ **GRACE**
 - ▶ Orbit/Gravity
 - ▶ General
 - ▶ **GGP**
 - ▶ General
 - ▶ **GGOS**
 - ▶ General
 - ▶ **GGSP**
 - ▶ General

[New Features and Documents](#) |

The [Product Type](#) description section is now available. New documents describing the ISDC are accessible in the "Document" section.
Read full article: 'New Features and Documents'

ISDC

Posted by: rit on Sep 22, 2006 - 03:46 PM

[New GRACE SDS Newsletter](#) |

The [GRACE Science System Monthly Report August 2006](#) prepared by Frank Flechtner (GFZ Potsdam) is available now.
Read full article: 'New GRACE SDS Newsletter'

Posted by: rit on Sep 15, 2006 - 02:26 PM

[New ISDC Collaboration Project](#) |

In cooperation with the [ICGEM](#) at the GFZ Potsdam a new ICGEM ISDC service will be established. Read full article: 'New ISDC Collaboration Project'

Posted by: rit on Aug 15, 2006 - 12:39 PM

Methodology

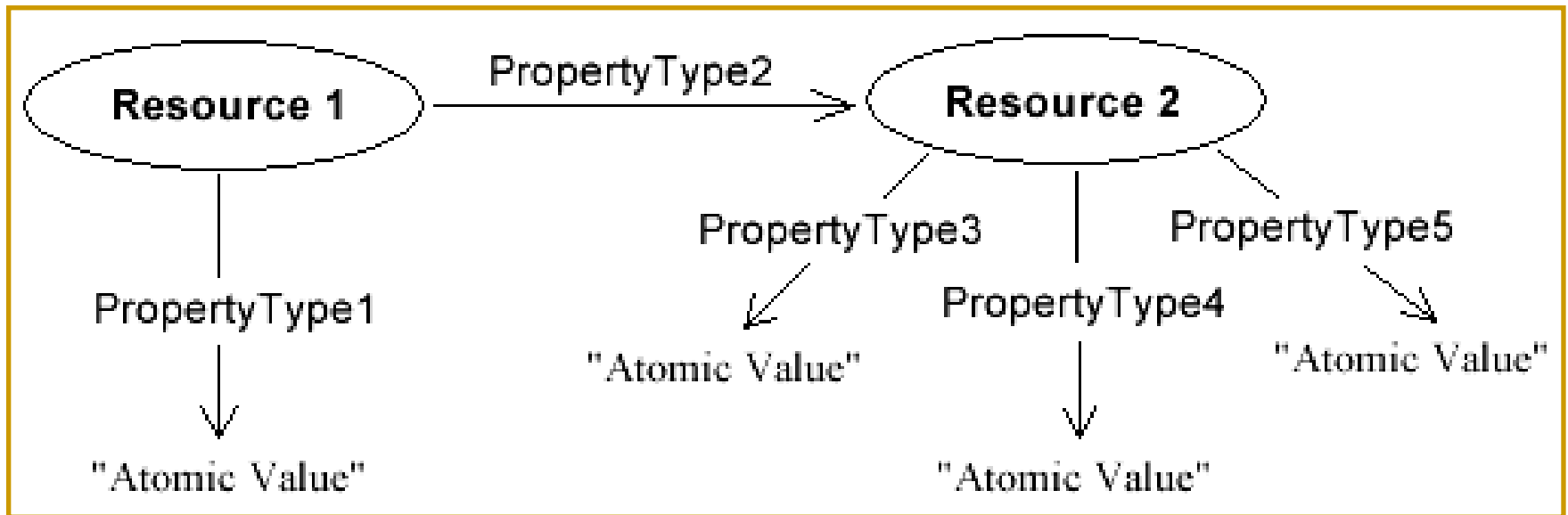
- Low functional capabilities for data processing and analysis
- Significant lack of datasets on Northern Asia and Siberia
- Universal solution: use of specialized web-sites and portals
 - Distributed dataset storage
 - Metadata for machine-processing description of datasets
 - Metadatabase
 - Set of tools for working with metadata
 - Online applications for environmental data processing

**Integrated information-computational system
based on web technologies**

Metadata standards

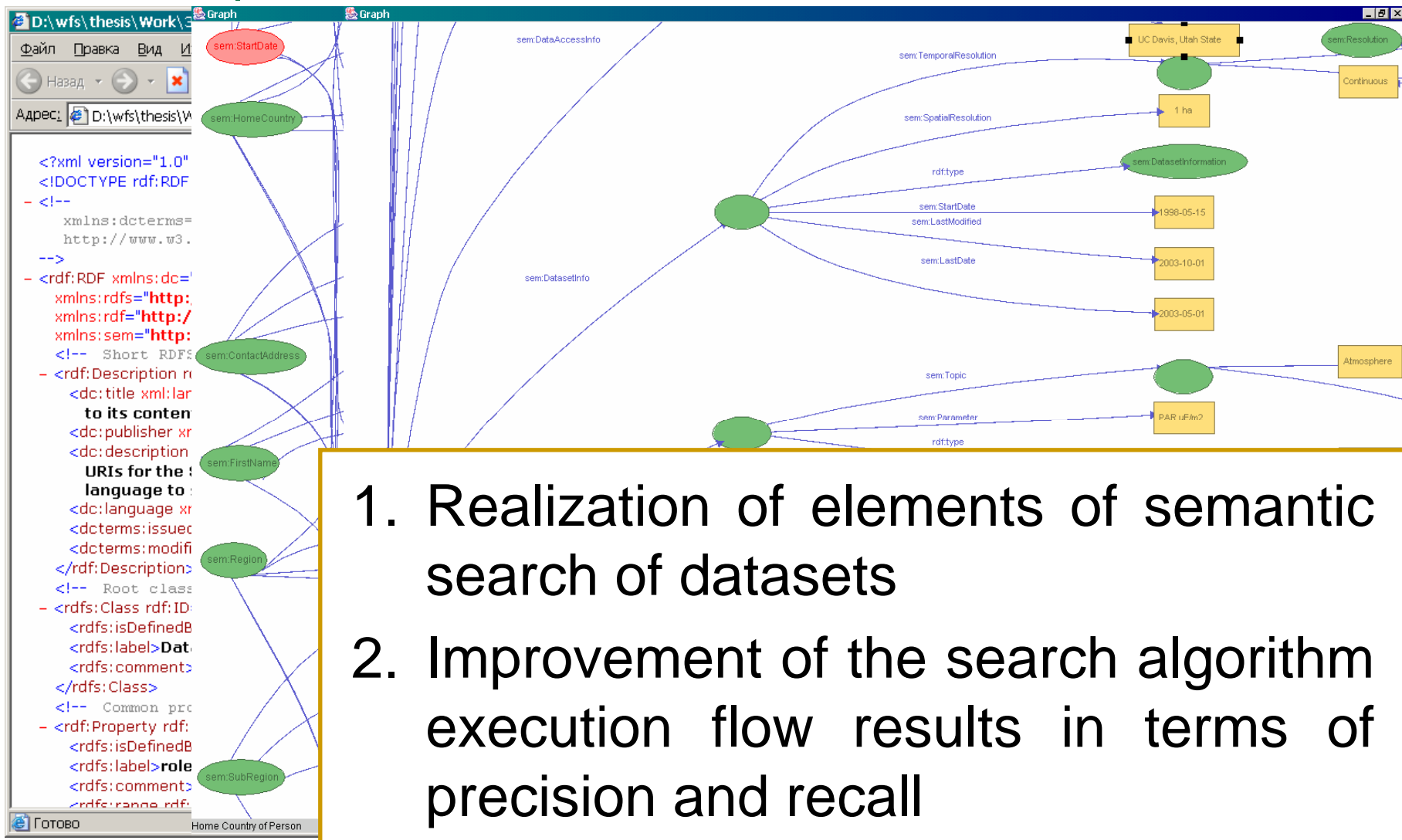
- Dublin Core Metadata Element Set (standard for cross-domain information resource description)
- EML, <http://knb.ecoinformatics.org/software/eml/>
- DIF, <http://gcmd.gsfc.nasa.gov/User/difguide/difman.html>
- NERIN XML Schema
- FGDC DTD, Content Standard for Digital Geospatial Metadata
- ISO 19115:2003 «Geographic information – Metadata»
- ISO 19139 «Geographic information – Metadata – XML schema implementation»
- SWEET ontologies (Semantic Web for Earth and Environmental Terminology), <http://sweet.jpl.nasa.gov/>

RDF Schema for metadata description



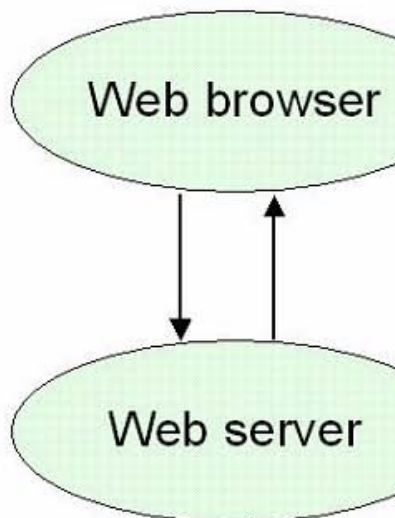
- Atomic values (string, numbers)
- Another resources with properties
- Displayed as a directed labeled graph (node and arc diagram)

RDF Schema for describing meteorology and climate datasets, version 0.1



1. Realization of elements of semantic search of datasets
2. Improvement of the search algorithm execution flow results in terms of precision and recall

General system model architecture



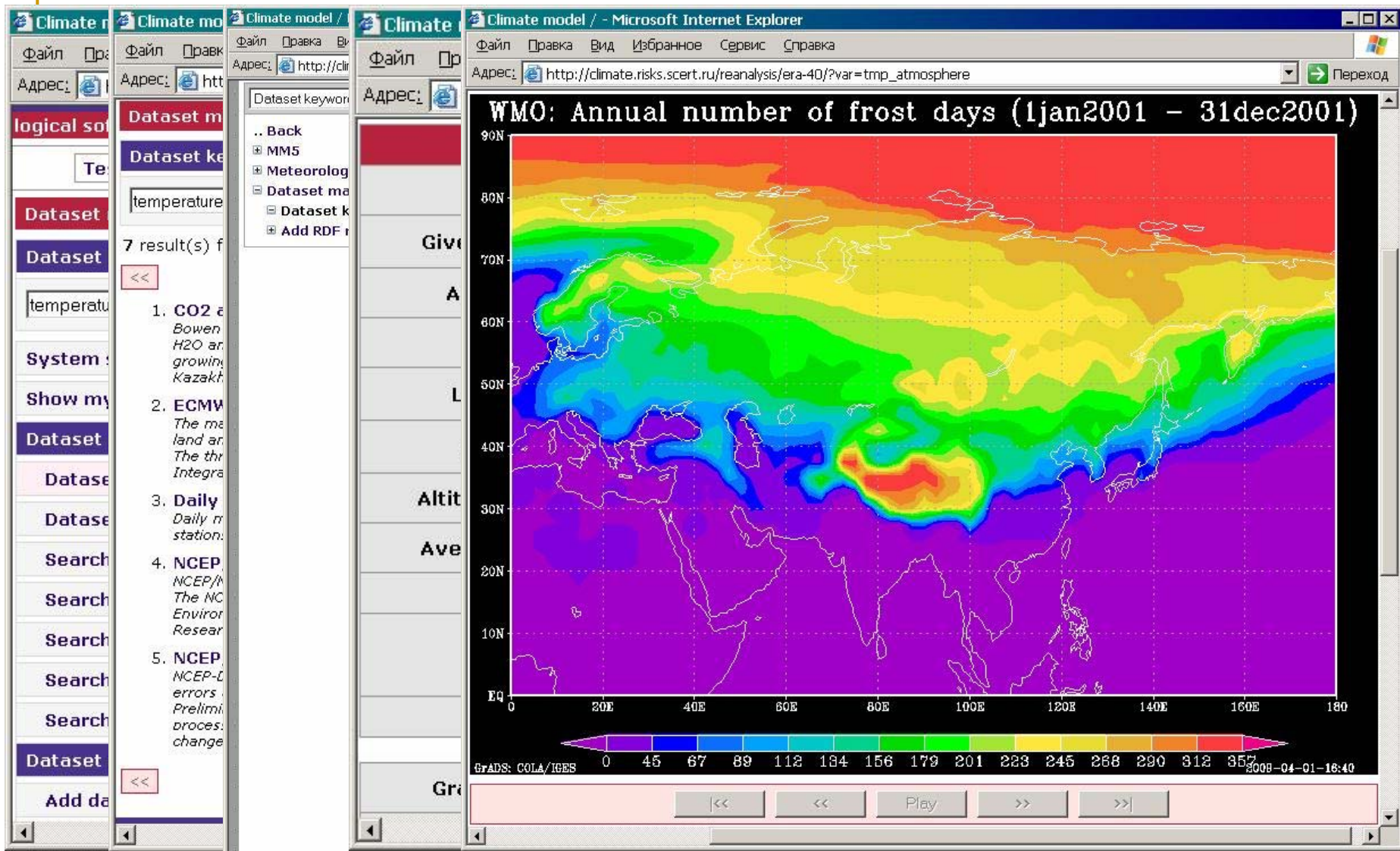
1. Dataset management functionality including SeRQL (Sesame RDF Query Language)-based semantic search
2. Statistical analysis and visualization of selected data archives

URL: <http://climate.risks.scert.ru/metadatabase/>

Software used:

- ATMOS web portal, <http://atmos.iao.ru/>
- Apache Tomcat, <http://tomcat.apache.org/>
- Sesame Server, <http://www.openrdf.org/>
- MySQL, <http://www.mysql.org/>

System model usage example



Statistical analysis operations

- Maximum, minimum, average, standard deviation values for parameter, time trends
- Correlation coefficient for an arbitrary pair of parameters
- Climate Change Indices approved by WMO
 - Number of frost/summer/icing days
 - Monthly maximum/minimum value of daily maximum/minimum temperature
 - Monthly maximum 1-day precipitation
 - Etc...

Conclusion

The proposed system architecture allows usage of advantages of the Semantic Web technologies and represents a step in the process of development of a distributed collaborative information-computational environment to support multidisciplinary investigations of Earth regional environment.

- Application in meteorological and climatology investigations
 - Providing access to data archives
 - Simplifying handling of huge arrays of spatially distributed meteorological data
-

Future plans

- Modifying RDF Schema to provide semantic references between resources
- Development of corresponding GUI for dataset annotation
- Statistical reports on dataset usage
- Extending functional capabilities of the web-system for data processing and visualization



Thank you for attention