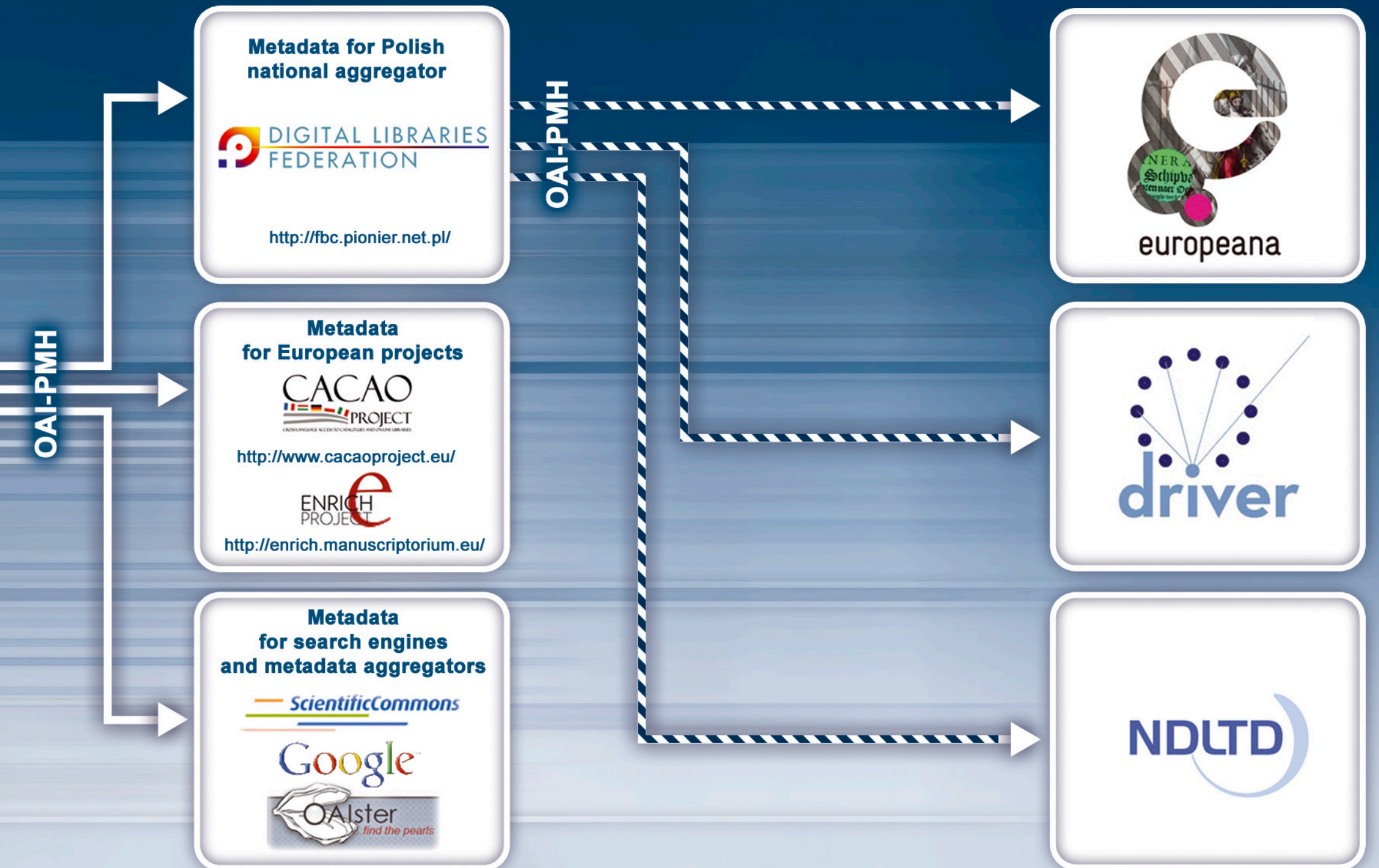
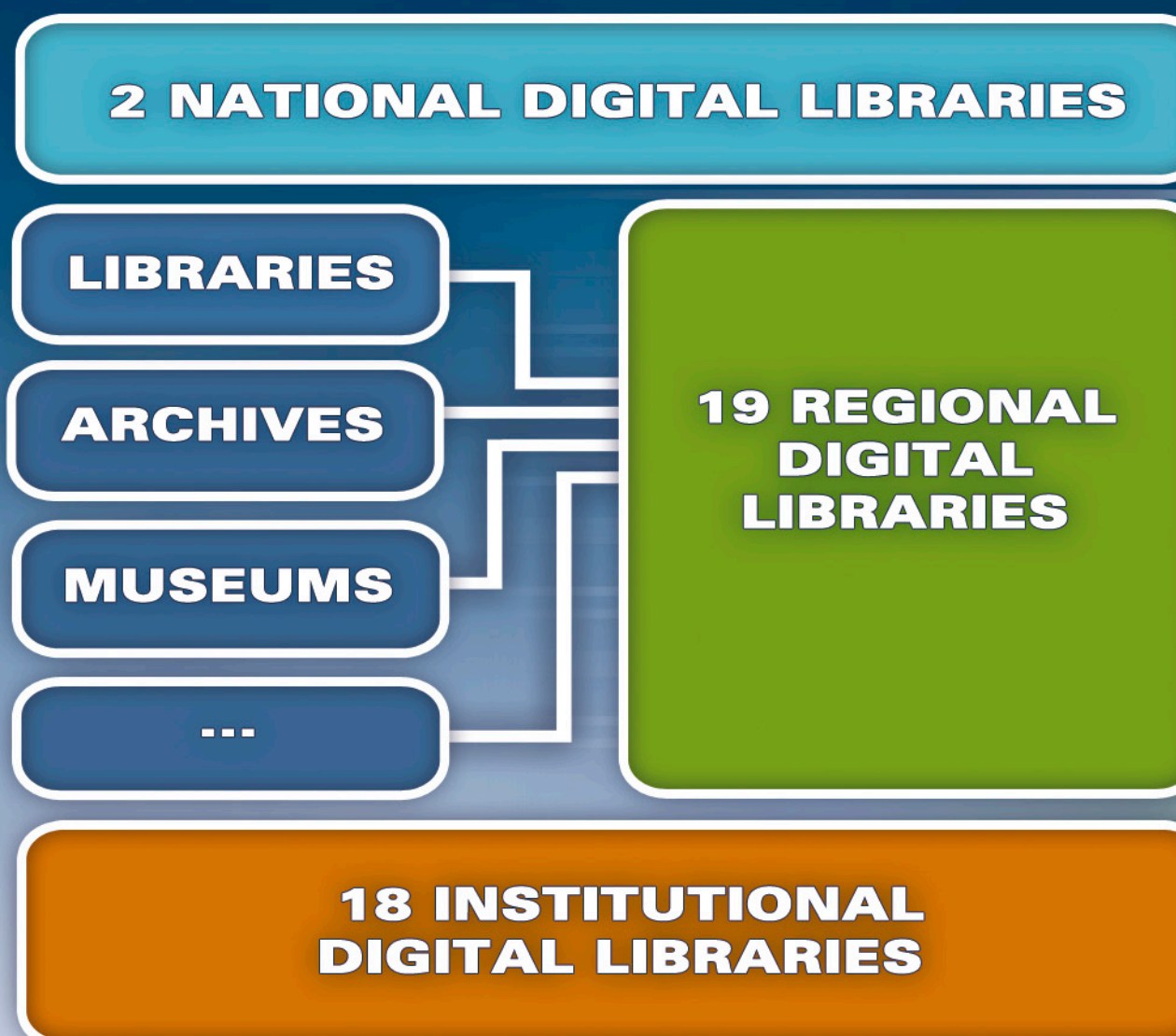


Federated regional and institutional digital libraries in Poland as a part of European data infrastructure

Adam Dudczak, Agnieszka Lewandowska, Marcin Werla (Poznań Supercomputing and Networking Center, Poznań, Poland)

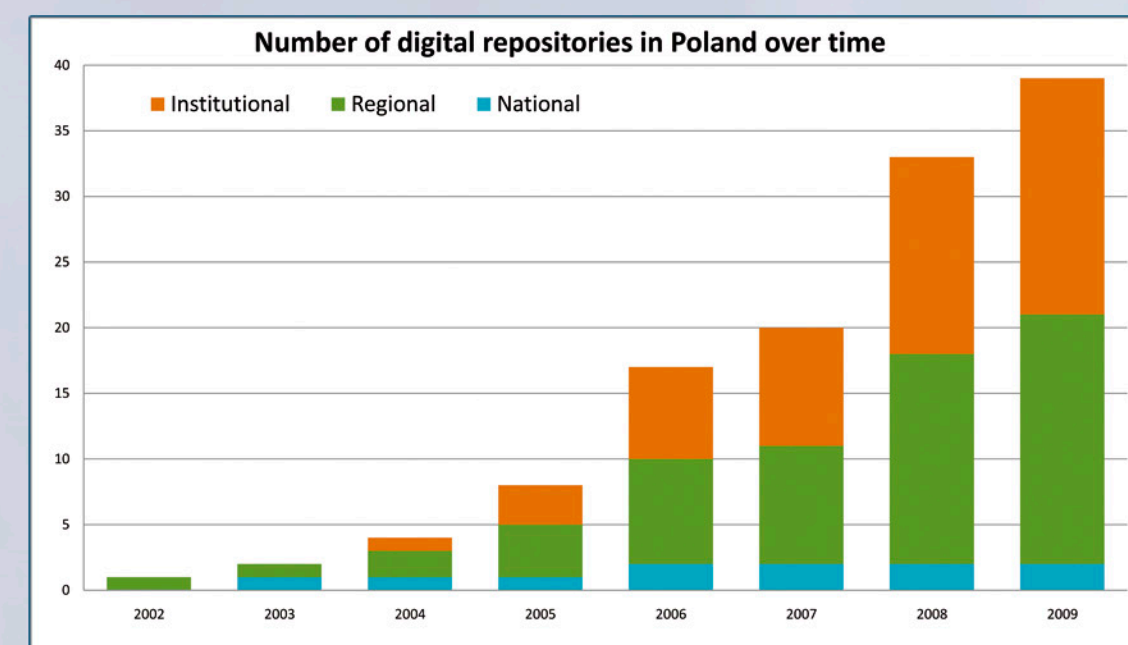


Over 300 institutions participate in building digital libraries in Poland



DEVELOPMENT OF DIGITAL LIBRARIES IN POLAND

Since 1999 Poznan Supercomputing and Networking Center (PSNC) has been developing the dLibra framework which aims to allow easy creation of distributed digital libraries (<http://dlibra.psnk.pl/>). In 2001 this software became a part of the Polish Optical Internet PIONIER programme. In October 2002, the first dLibra-based regional digital library, the Digital Library of the Wielkopolska (<http://www.wbc.poznan.pl/>), was made publicly available. Currently this library holds more than 80 000 digital objects and is the largest digital library in Poland. Nowadays Polish platform of distributed digital libraries in the PIONIER network, consists of **39 publicly available digital libraries** (19 regional, 18 institutional and 2 national). Together they give access to over **250 000 various multimedia objects**, including contents from hundreds of libraries, archives and museums.



Those digital libraries may be divided into three types, based on their character and the number of participating institutions. An **institutional digital library** is usually maintained by one institution (e.g. university) and gives access to digital materials related to activities led by the institution. A **regional digital library** is usually supported by a number of institutions (scientific and cultural institutions, like libraries, museums and archives, etc.) from a given region and gives access to digital materials related to the region cultural and scientific heritage. Technical maintenance is usually provided by a regional computing center. Finally, a **national digital library** is maintained by a national institution and is in some way unique for the entire country.

PIONIER: POLISH OPTICAL INTERNET - ADVANCED APPLICATIONS, SERVICES AND TECHNOLOGIES FOR INFORMATION SOCIETY

Polish Optical Internet PIONIER is a new generation national research and education optical network. This initiative has triggered many supporting activities, aiming at building the base mechanisms of information society and providing Poland for equal partnership opportunities with other countries in: science, education, health care, natural environment, government and local administration, industry and services. Below one can find a list of Polish digital libraries deployed in PIONIER network.

National digital libraries:

- National Digital Library "Polona" (Warszawa)
- Polish Internet Library - PBI (Warszawa)

Regional digital libraries:

- Baltic Digital Library (Slupsk)
- Digital Land of Sieradz (Sieradz)
- Lower Silesian Digital Library (Wroclaw)
- Digital Library of Elblag (Elblag)
- Digital Library of Jelenia Gora (Jelenia Gora)
- Kujawsko-Pomorska Digital Library (Torun)
- Malopolska Digital Library (Krakow)
- Digital Library of Nowa Huta (Krakow)
- Podkarpacka Digital Library (Rzeszow)
- Podlaska Digital Library (Bialystok)
- Digital Library of Radom (Radom)
- Digital Library - Collection of the Lodz Region (Lodz)
- Sanok Digital Library (Sanok)
- Silesian Digital Library (Katowice)
- Swietokrzyska Digital Library (Kielce)
- Digital Library of Wejherowo (Wejherowo)
- Digital Library of the Wielkopolska (Poznan)
- West Pomeranian Digital Library "Pomerania" (Szczecin)
- Digital Library of Zielona Gora (Zielona Gora)

Institutional digital libraries:

- Academic Digital Library - CRACOW (Krakow)
- Digital Library of Book Studies (Warszawa)
- Digital Library of Cracow University of Technology (Krakow)
- CODN Digital Library (Warszawa)
- Lublin University of Technology Digital Library (Warszawa)
- Technical University of Lodz eBIPoL Digital Library (Lodz)
- Warsaw University of Technology Digital Library (Warszawa)
- Digital Library of the University of Lodz (Lodz)
- Digital Library of University of Warmia and Mazury (Olsztyn)
- Digital Library of Wroclaw University (Wroclaw)
- MCS University Digital Library (Lublin)
- Silesian University of Technology Digital Library (Gliwice)
- e-BUW - the digital library of the University of Warsaw (Warszawa)
- Swietokrzyska Digital Library (Kielce)
- Genealogical Digital Library (Warszawa)
- Digital Library of the Gdansk Library of the Polish Academy of Sciences (Gdansk)
- Digital Repository of ICM (Warszawa)
- FIDES Digital Library (Warszawa)
- Pedagogical Digital Library (Krakow)

METADATA AGGREGATION

Metadata of digital objects available in Polish digital libraries is used for various purposes by the external network services such as **search engines** (e.g. Google, Yahoo), **metadata aggregators** (e.g. PIONIER DLF, OAIster) or **domain/thematic services** (e.g. Manuscriptorium, CACAO project). While the search engines and metadata aggregators harvest all available metadata, the domain/thematic services use **selective harvesting** for their purposes.

PIONIER Network Digital Libraries Federation (PIONIER DLF) is a Polish metadata aggregator, which aggregates metadata from all OAI-PMH compliant Polish digital libraries. The mission of this service is to:

- facilitate the use of resources of Polish digital libraries and repositories,
- increase the visibility of Polish digital resources in the Internet,
- give Internet users access to new, advanced network services based on the resources of Polish digital libraries and repositories.

This mission is realized by constant development of the PIONIER DLF functionality and by the popularization of the PIONIER DLF. The harvested metadata of available digital objects is used in several ways. First of all, it is indexed and then it may be searched with simple and advanced queries. This gives the possibility to search the metadata of all Polish digital resources in one service. Secondly, it is used for the OAI identifier resolution. Users may use the following URL syntax:

<http://fbc.pionier.net.pl/id/<oaiidentifier>>

which results in redirection to the current localization of a digital object identified by a given OAI identifier. Thirdly, the harvested data is used to generate several statistics and comparisons (e.g. the percentage of digital objects types harvested by the PIONIER DLF).

All dLibra-based digital libraries have the possibility to store the data about their digitization plans, which is also exposed via the OAI-PMH protocol. This data is used by PIONIER DLF to prepare a report containing descriptions of all objects currently planned for digitization in Poland (about 3 500 objects in average). Additionally this data and the data about already available digital objects is used in the duplicates detection mechanism. This mechanism is used to generate a report with digital objects coming from different digital libraries, but having very similar metadata (title, creator, publication date etc.). Such digital objects may in fact represent the same physical object digitized twice in two separate institutions. Currently this report contains almost 400 potentially duplicated digital objects. The duplicate detection mechanism is also exposed as REST web service for external use.

The PIONIER DLF search plug-in is one of the preinstalled plug-ins in the Polish localization of the Firefox 2 and 3 web browsers.

EXTENDED HARVESTING CAPABILITIES

Besides of being an OAI-PMH service provider, the PIONIER DLF acts also as the fully OAI-PMH compliant data provider, with extended selective harvesting functionality. This allows external projects (e.g. Europeana, DRIVER) to perform selective harvesting on the national-level. Selective harvesting is based on the concept of dynamic sets.

Dynamic set is a set of items, which is not defined in repository prior to the client request. The set membership criterion is defined by the set specification in the client request. In case of vertical/thematic harvesters (like DRIVER, NDLD) the use of the OAI-PMH extension based on dynamic sets decreases the number of records transferred from the repository to the harvester. Dynamic sets can be harvested by any OAI-compliant harvester, no software changes are required. Criteria for set membership are defined by the harvester, the only place where criteria can be placed without adding additional parameters to the request is the set specification. So the OAI request with dynamic set specification can look like this:

`verb=ListRecords&metadataPrefix=oai_dc&set=SomeSet:EncodedCriteria`

which means that returned items should be from SomeSet and additionally the items should match the EncodedCriteria. The criteria are encoded in the CQL language. CQL is a query language designed for various information retrieval systems. To conform to the restrictions of the OAI set specification, the CQL-based dynamic OAI set specifications should be URL-encoded (e.g. `dc.creator%3D%22Albert%20Einstein%22`).



www: <http://www.fbc.pionier.net.pl>
e-mail: fbc@lists.man.poznan.pl

POZNAN SUPERCOMPUTING AND NETWORKING CENTER
Noskowskiego 12/14, 61-704 Poznan, Poland
<http://www.man.poznan.pl>