



NTK

50°6'14.083"N, 14°23'26.365"E

Národní technická knihovna
National Technical Library

CDS Invenio

Software solution for National
Repository of Grey Literature (NRGL)

Tomáš Müller

8.12. 2010

Seminář ke zpřístupňování
šedé literatury 2010

Grey literature

210 mm

- **Definition:** "Information produced on all levels of government, academics, business and industry in electronic and print formats not controlled by commercial publishing i.e. where publishing is not the primary activity of the producing body." (Luxembourg, 1997 - Expanded in New York, 2004)
- E.g. Research reports, theses, proceedings, e-mails, ...
- Might contain valuable information
- Hard to get

210 mm

- Central repository
- Harvesting metadata and full texts
- Cooperation with organizations
- Hosting seminars
- Joining international projects

CDS Invenio

210 mm

- CERN
- Document management system
- Platform: Unix-like
- Freeware, open source
- Python, MySQL,...
- <http://invenio.ntkcz.cz>

INVENIO 

CDS Invenio - advantages

210 mm

- Speed optimization
- Cache, cache, cache!
- Indexes
- Modular architecture
- Expandable

CDS Invenio - disadvantages

210 mm

- Version 0.99.1 (bugs)
- Difficult administration
- Inconsistent and incomplete documentation

Invenio installation

210 mm

- Virtual environment (Virtualbox)
- OS: Debian
- Get rid of platform dependence
- Possibility to export the virtual machine
- Invenio installation is long and tedious



Invenio – document structure

210 mm

- Division into collections
- Two ways of division 1. by document type, 2. by institution
- Document types: theses, Reports, Copyrighted work, ...

Invenio – acquiring records

210 mm

- 1. Direct insertion via web interface
- 2. Harvesting using OAI-PMH protocol
([example](#), [example2](#))
- 3. By e-mail (we don't use this)

Record – MARC21

210 mm

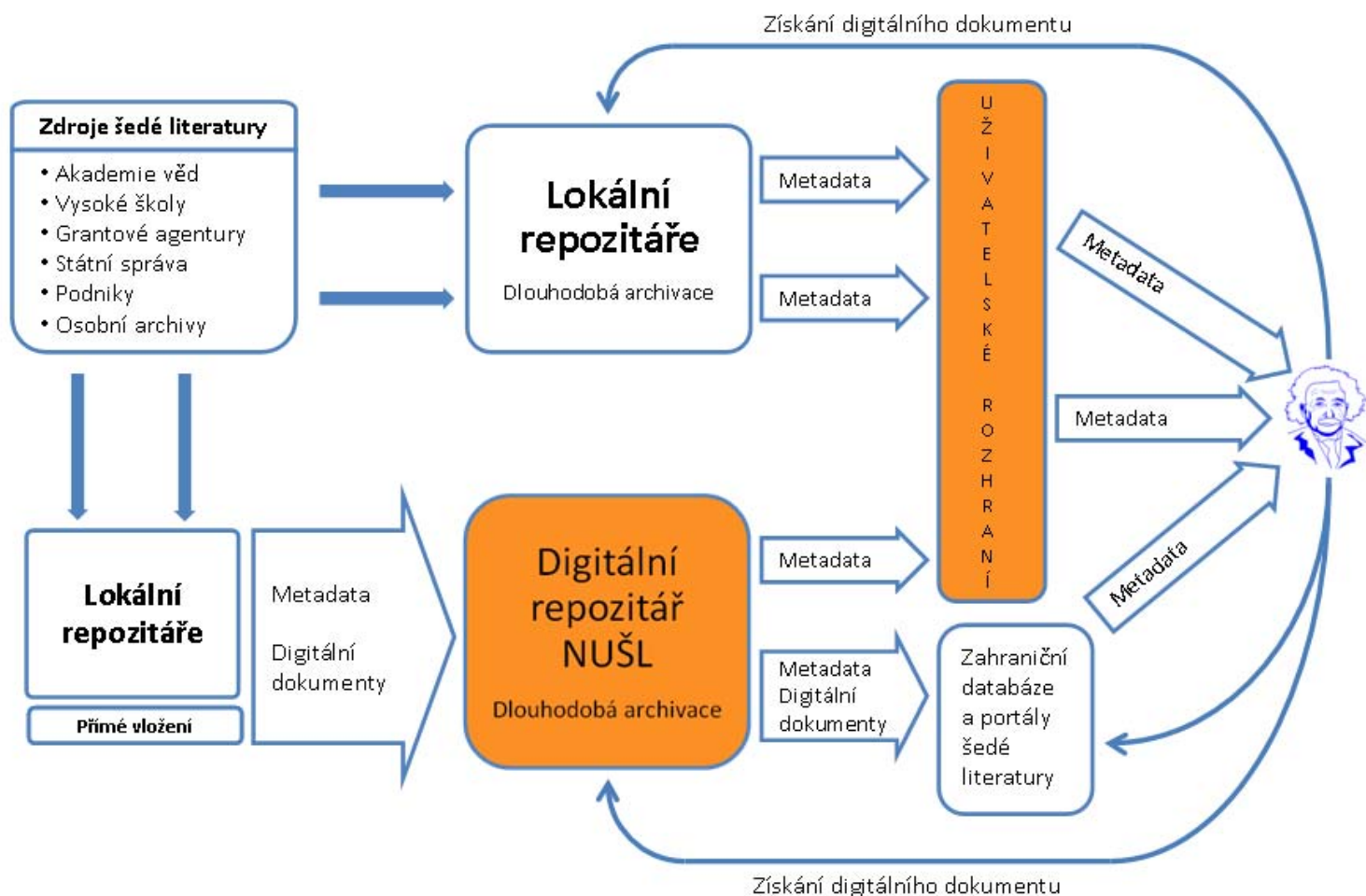
```
000000075 001__ 75
000000075 005__ 20100907085626.0
000000075 008__ 100907n|||||||xx#||||#|||z
000000075 0248_ $$anusl-75
000000075 04107 $$acze
000000075 046__ $$k2010
000000075 046__ $$j2010
000000075 24510 $$aNárodní úložiště šedé lit
000000075 24633 $$a$$b
000000075 24630 $$n$$p
000000075 340__ $$aapplication/pdf
000000075 506__ $$aPublic
000000075 520__ $$aCílem příspěvku je předst
000000075 540__ $$aText je chráněný podle au
000000075 650_7 $$2PSH$$aobecnosti
000000075 6531_ $$adigitální
000000075 6531_ $$arepozitář
000000075 6531_ $$aúložiště
000000075 6531_ $$anárodní
000000075 6531_ $$ašedá
000000075 6531_ $$aliteratura
000000075 6531_ $$aNárodní
000000075 6531_ $$atechnická
```

```
<collection xmlns="http://www.loc.gov/MARC21" >
<record>
<controlfield tag="001">75</controlfield>
<controlfield tag="005">20100907085626.0</controlfield>
<controlfield tag="008">100907n|||||||xx#||||#|||z</controlfield>
<datafield tag="024" ind1="8" ind2=" " >
<subfield code="a">nusl-75</subfield>
</datafield>
<datafield tag="041" ind1="0" ind2="7">
<subfield code="a">cze</subfield>
</datafield>
<datafield tag="046" ind1=" " ind2=" " >
<subfield code="k">2010</subfield>
</datafield>
<datafield tag="046" ind1=" " ind2=" " >
<subfield code="j">2010</subfield>
</datafield>
<datafield tag="245" ind1="0" ind2="0">
<subfield code="a">Národní úložiště šedé lit</subfield>
<subfield code="b"></subfield>
</datafield>
<datafield tag="246" ind1="3" ind2="3">
<subfield code="a"></subfield>
</datafield>
```

210 mm

- <http://www.nusl.cz>
- User search interface
- Connected to Invenio and other repositories

FAST – scheme



Current work

210 mm

- Migrating Invenio into new virtual environment
- Harvesting sources
- Joining international projects
- Enhancing modules

The end

210 mm

Thank you for your attention