



Enabling Grids for E-science

Úvod do gridového počítania

Miroslav Dobrucký
Ústav informatiky
Slovenská akadémia vied
Bratislava

www.eu-egee.org



- **Čo je to gridové počítanie?**
- **Charakteristiky gridu**
- **Aplikácie**
- **Gridové projekty**

Vedci a výskumníci vykonávajú svoje aktivity v kooperácii v rámci celého sveta, komunikujú medzi sebou a zdieľajú spolu údaje.

Grid: zosieťované centrá na spracovanie údajov a "middleware" ako "lepidlo" na spojenie zdrojov.



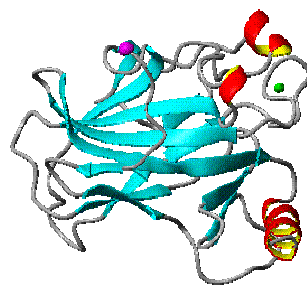
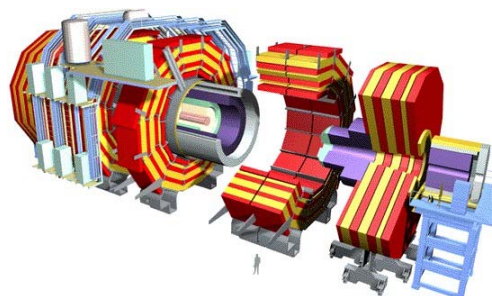
Vedecké nástroje a experimenty poskytujú obrovské množstvo údajov.

- **Virtuálna organizácia je:**
 - Keď ľudia z rôznych inštitúcií pracujú na riešení spoločných úloh
 - A zdieľajú distribuované výpočtové a údajové zdroje
- **Gridová infraštruktúra podporuje virtuálne organizácie**

“Grid computing is coordinated resource sharing and problem solving in dynamic, multi-institutional virtual organizations”
(I.Foster)

- **Súčasné distribuované aplikácie:**
 - Smerujú k *špecializovným systémom*
 - Sú zamerané na jeden účel alebo skupinu používateľov
- **Gridy idú ďalej a berú do úvahy:**
 - Rôzne druhy zdrojov
 - Nie vždy rovnaký hardvér, údaje a aplikácie
 - Rôzne druhy *interakcií*
 - Skupiny používateľov alebo aplikácie chcú interagovať s gridom rôznymi spôsobmi
 - *Dynamická povaha*
 - Zdroje a používatelia sú pridávaní/odoberaní/menení často

- **Neustále zvyšovanie výkonu mikroprocesorov**
 - Možno postaviť multi-gigaflop-ový systém za menej než 30tis.Sk
- **Túžba posunúť hranice vedeckého bádania pomocou simulácií a analýzy – e-veda**



- Čo je to gridové počítanie?
- **Charakteristiky gridu**
- **Aplikácie**
- **Gridové projekty**



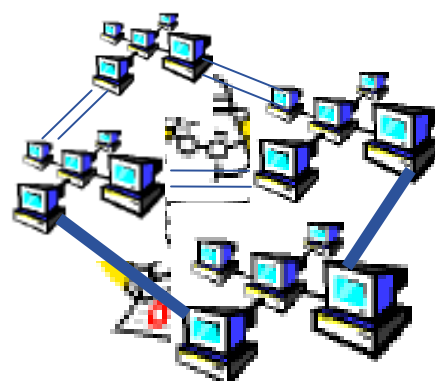
Štandardy

Veľké množstvo zdrojov

Vlastnené navzájom si nedôverujúcimi organizáciami a ľuďmi

Rôzne požiadavky na bezpečnosť a politiku

Potenciálne poruchové zdroje



Prepojené heterogénnymi, viac-úrovňovými sieťami

Rôzne politiky manažmentu zdrojov

Geograficky oddelené

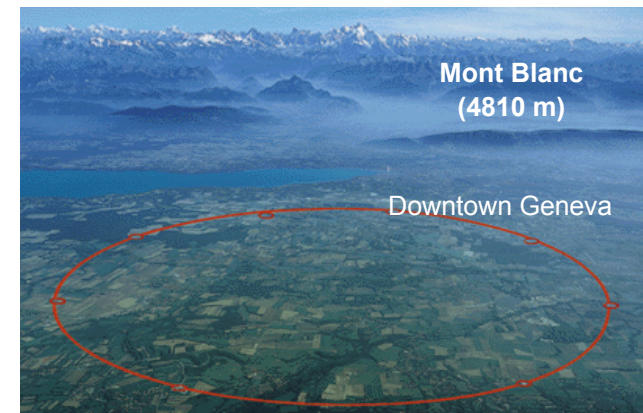
Zdroje sú heterogénne

- Čo je to gridové počítanie?
- Charakteristiky gridu
- **Aplikácie**
- **Gridové projekty**

- **Medické/zdravotnícke** (*snímkovanie, diagnostika a liečenie*)
- **Bioinformatické** (*skúmanie ľudského genómu aby sme porozumeli genetickým chorobám*)
- **Nanotechnológie** (*návrh nových materiálov na molekulárnej úrovni*)
- **Strojárstvo** (*optimalizácia návrhu, simulácie, analýza porúch a vzdialená manipulácia a riadenie*)
- **Prírodné zdroje a životné prostredie** (*predpoveď počasia, pozorovanie zeme, modelovanie a predikcia zložitých systémov*)



- **The Large Hadron Collider (LHC)**
 - doteraz postavený **najvýkonnejší nástroj** na vyšetovanie elementárnych častíc
- **Údajová výzva:**
 - **10 Petabajtov/rok údajov !!!**
 - **20 miliónov CD-čiek každým rokom!**
- **Simulácia, rekonštrukcia, analýza:**
 - Údaje z LHC vyžadujú výpočtový výkon porovnateľný s **~100,000 najrýchlejšími súčasnými PC procesormi!**



- Čo je to gridové počítanie?
- Charakteristiky gridu
- Aplikácie
- **Gridové projekty**

Mnohé snahy vyvíjať Grid — na celom svete

- 
- NASA Information Power Grid
 - DOE Science Grid
 - NSF National Virtual Observatory
 - NSF GriPhyN
 - DOE Particle Physics Data Grid
 - NSF TeraGrid
 - DOE ASCI Grid
 - DOE Earth Systems Grid
 - DARPA CoABS Grid
 - NEESGrid
 - DOH BIRN
 - NSF iVDGL
 - DataGrid (CERN, ...)
 - EuroGrid (Unicore)
 - DataTag (CERN,...)
 - Astrophysical Virtual Observatory
 - GRIP (Globus/Unicore)
 - GRIA (Industrial applications)
 - GridLab (Cactus Toolkit)
 - CrossGrid (Infrastructure Components)
 - EGSO (Solar Physics)
 - UK – OGSA-DAI, RealityGrid, GeoDise, Comb-e-Chem, DiscoveryNet, DAME, AstroGrid, GridPP, MyGrid, GOLD, eDiamond, Integrative Biology, ...
 - Netherlands – VLAM, PolderGrid
 - Germany – UNICORE, Grid proposal
 - France – Grid funding approved
 - Italy – INFN Grid
 - Eire – Grid proposals
 - Switzerland - Network/Grid proposal
 - Hungary – DemoGrid, Grid proposal
 - Norway, Sweden - NorduGrid

European DataGrid (EDG)

www.edg.org



LHC Computing GRID (LCG)

cern.ch/lcg



CrossGRID

www.crossgrid.org



DataTAG

www.datatag.org



GridLab

www.gridlab.org



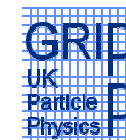
EUROGRID

www.eurogrid.org

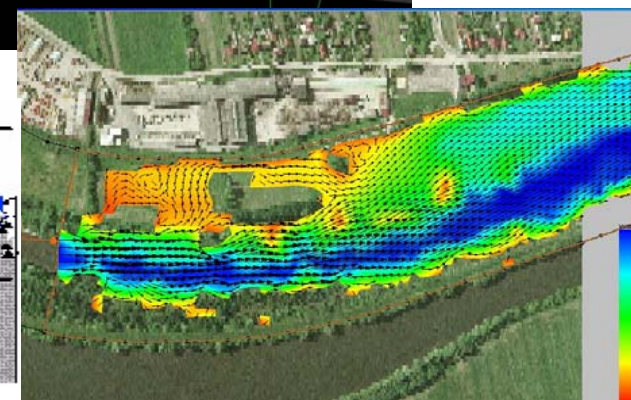
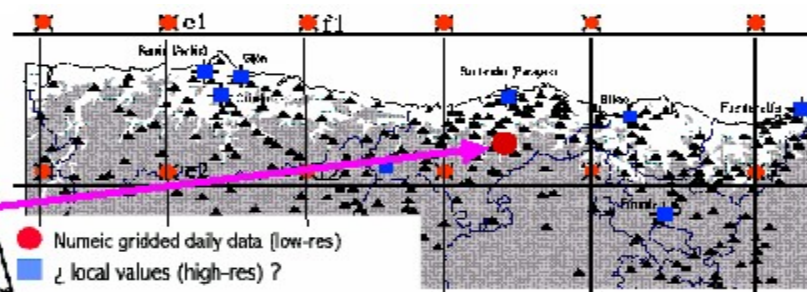
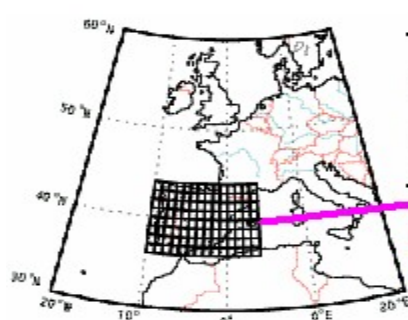
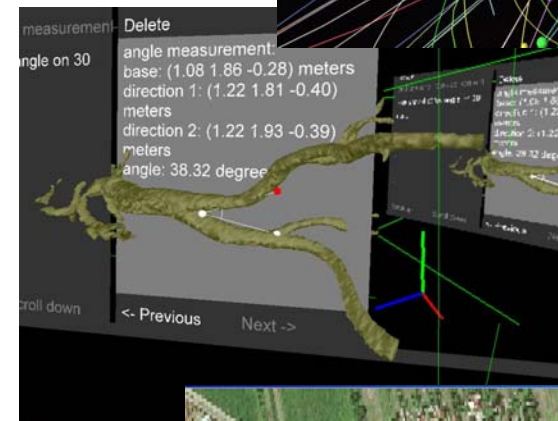
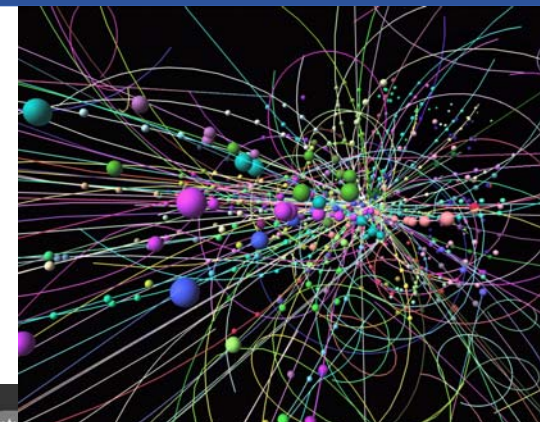


Európske národné projekty:

- INFN GRID,
- UK e-Science Programme,
- NorduGrid



- 1. Interaktívna biomedická simulácia a vizualizácia
- 2. Podporný systém pre protipovodňový krízový štáb
- 3. Distribuovaná analýza údajov fyziky vysokých energií
- 4. Predpoveď počasia a modelovanie znečistenia ovzdušia



Video
prenos

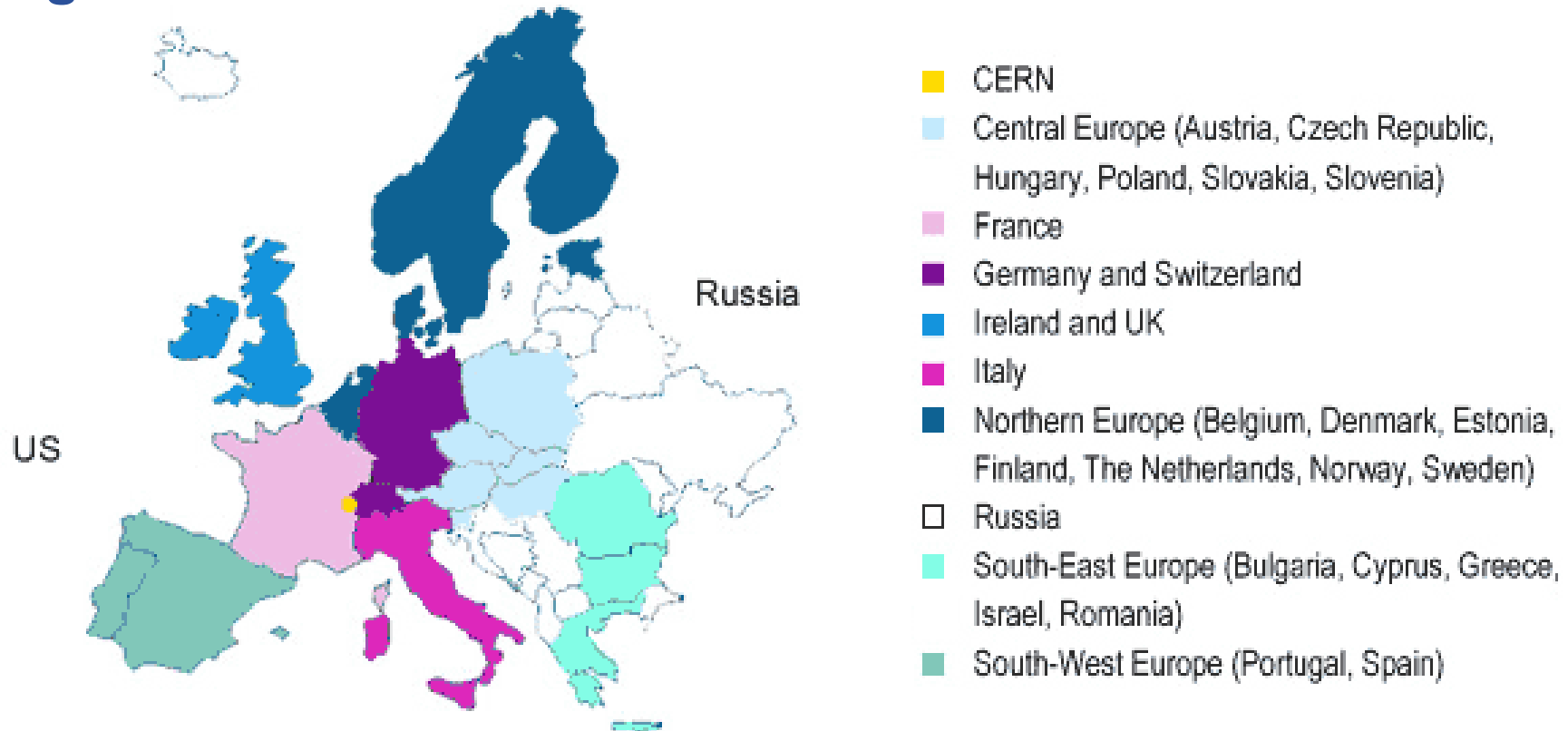


Vizualizácia

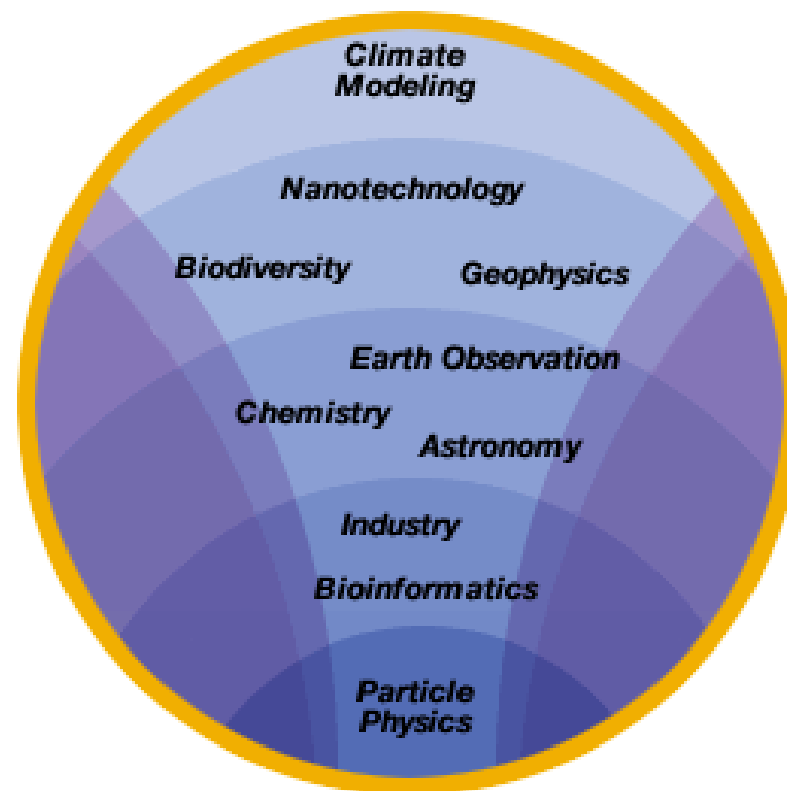
Mikrofóny

Kamery

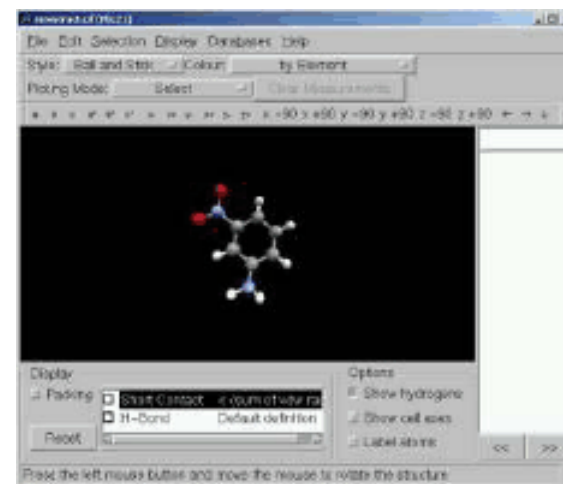
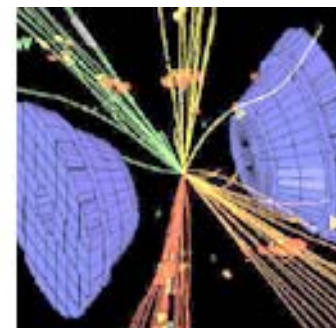
- Spoločné využívanie národných zdrojov za účelom dosiahnutia vyššieho úžitku pre Európu
- 70 inštitúcií z 27 krajín, združených v regionálnych gridoch



- **Rozsah projektu EGEE: plná podpora pre akademické aplikácie (a taktiež otvorený pre priemyselný a socio-ekonomický svet)**
- **Hlavné kritérium úspechu EGEE projektu: koľko spokojných používateľov z koľkých rôznych oblastí ?**
- **5000 používateľov (3000 po 2 rokoch) aspoň z 5-tich disciplín**
- **Dve vybrané pilotné aplikácie budú určovať implementáciu a zhodnotia výkonnosť a funkčnosť vyvíjanej infraštruktúry: Fyzika & Bioinformatika**



- **Fyzika vysokých energií (High Energy Physics)** s výpočtovým gridom LHC (www.cern.ch/lcg) využíva gridovú infraštruktúru na ukladanie a analýzu petabajtov (10^{15} bajtov) reálnych a simulovaných údajov. LCG projekt je hlavným zdrojom požiadaviek a tvrdých termínov, ktoré sa nedajú konvenčne riešiť.
- **V biomedicíne** mnohé komunity čelia rovnako skľučujúcim výzvam zvládnuť záplavu bioinformatických a zdravotníckych údajov. Potrebujú mať prístup k veľkým a distribuovaným nehomogénnym údajom a majú dôležitú požiadavku na okamžité výkonné počítanie.



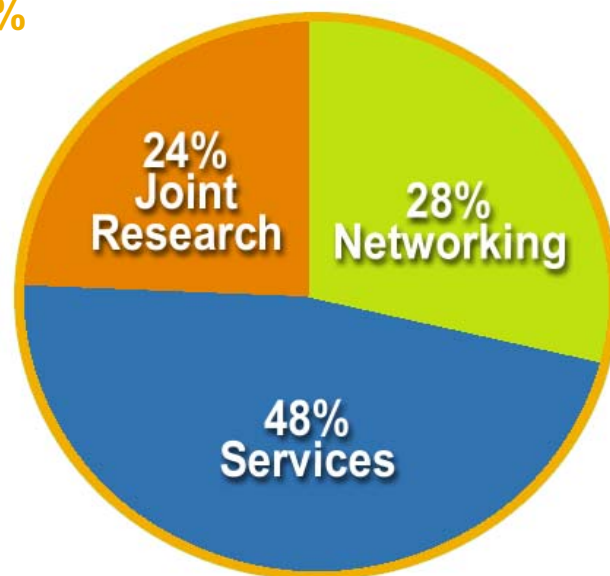
24% Združený výskum

JRA1: Príprava middleware a integrácia – 17%

JRA2: Zabezpečenie kvality – 1.5%

JRA3: Bezpečnosť – 3%

JRA4: Vývoj sieťových služieb - 2.5%



48% Služby

SA1: Prevádzka gridu, podpora a správa

SA2: Zabezpečovanie sieťovej infraštruktúry

28% Sieťové aktivity

NA1: Manažment

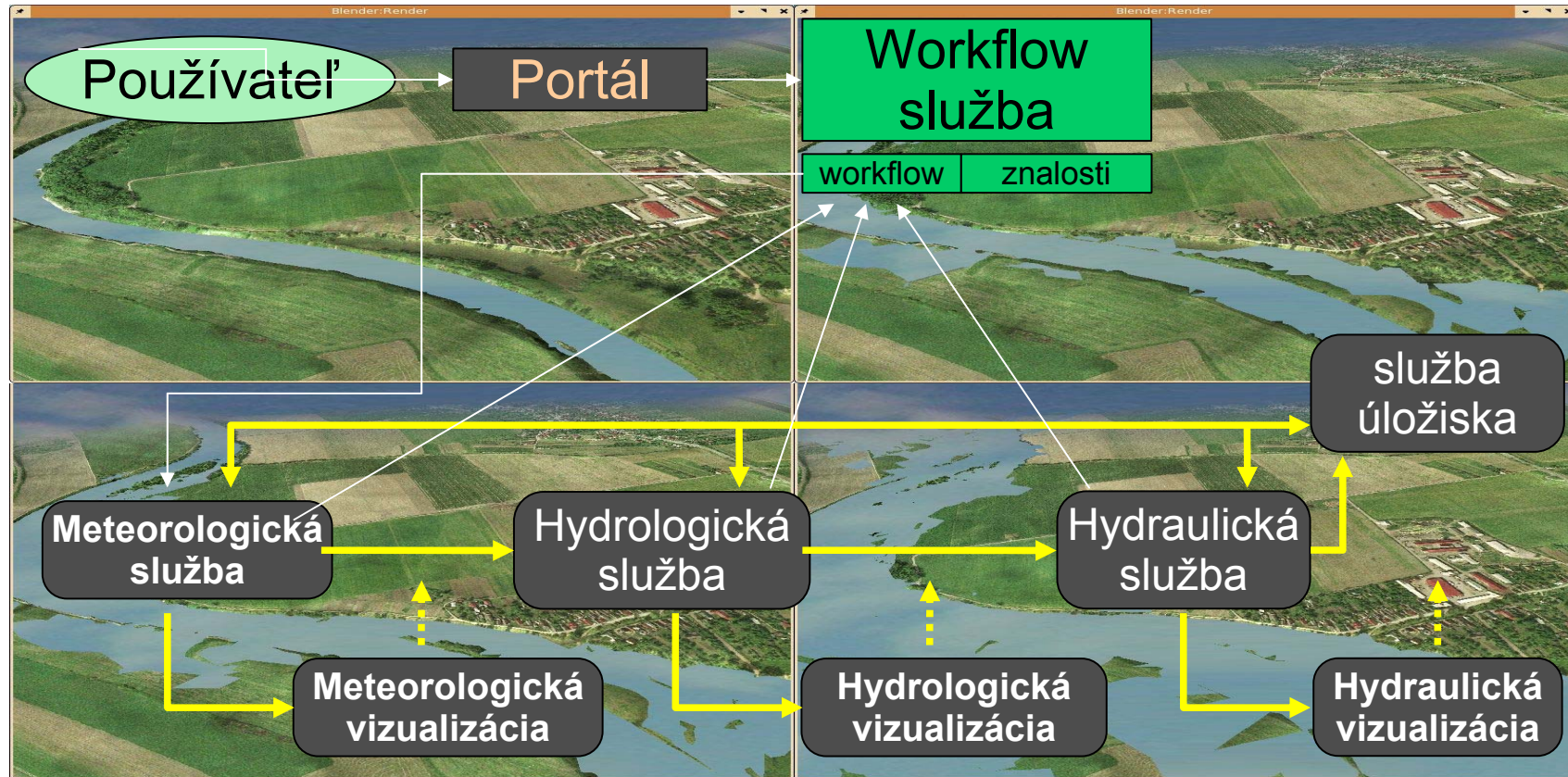
NA2: Disseminácia a propagácia

NA3: Trénovanie používateľov a vzdelávanie

NA4: Identifikácia aplikácií a ich podpora

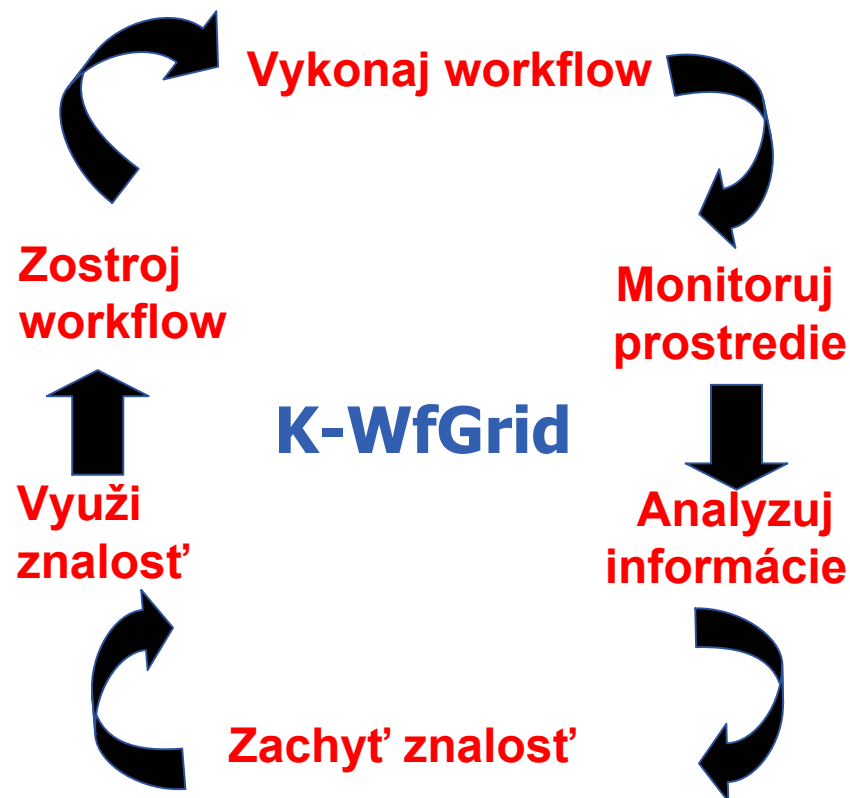
NA5: Politika a medzinárodná spolupráca

Dôraz v EGEE je na prevádzke produkčného gridu a na podpore koncových používateľov



Aplikácie sú zložité a dynamicky konštruované zo služieb.
 Súčasné riešenie je závislé na človekovi ako na zdroji znalostí.

- Integrovať gridové služby do koherentného aplikačného scenáru
- Umožniť automatickú tvorbu a opakované používanie toku prác (workflow) so znalosťou pozbieranou počas činnosti
- Monitorovanie gridu na získanie znalostí o službách za účelom poskytnutia pridanej hodnoty koncovému používateľovi



Technológie: gridová architektúra orientovaná na služby, softvéroví agenti, ontológie, dynamická inštrumentácia

Ďakujem za pozornosť

egee.ui@sav.sk

Miroslav Dobrucký

Ústav informatiky

Slovenská akadémia vied

Bratislava

www.eu-egee.org