CLARIN Issues

Peter Wittenburg
MPI for Psycholinguistics
Nijmegen, NL

What’s CLARIN

- one of the successful ESFRI proposals for research infrastructures

- mission
  - domain of language resources and technology is highly fragmented
  - little is visible, little fits together
  - CLARIN wants to build an integrated and interoperable landscape of LRT and offer easy to use LRT services to interested researchers

- state
  - > 170 member institutions from 32 “EU” countries
  - substantial EC funding
  - also substantial funding from various national agencies
  - some commitments already longer than 2010
  - Executive Board (8 experts) is leading the work

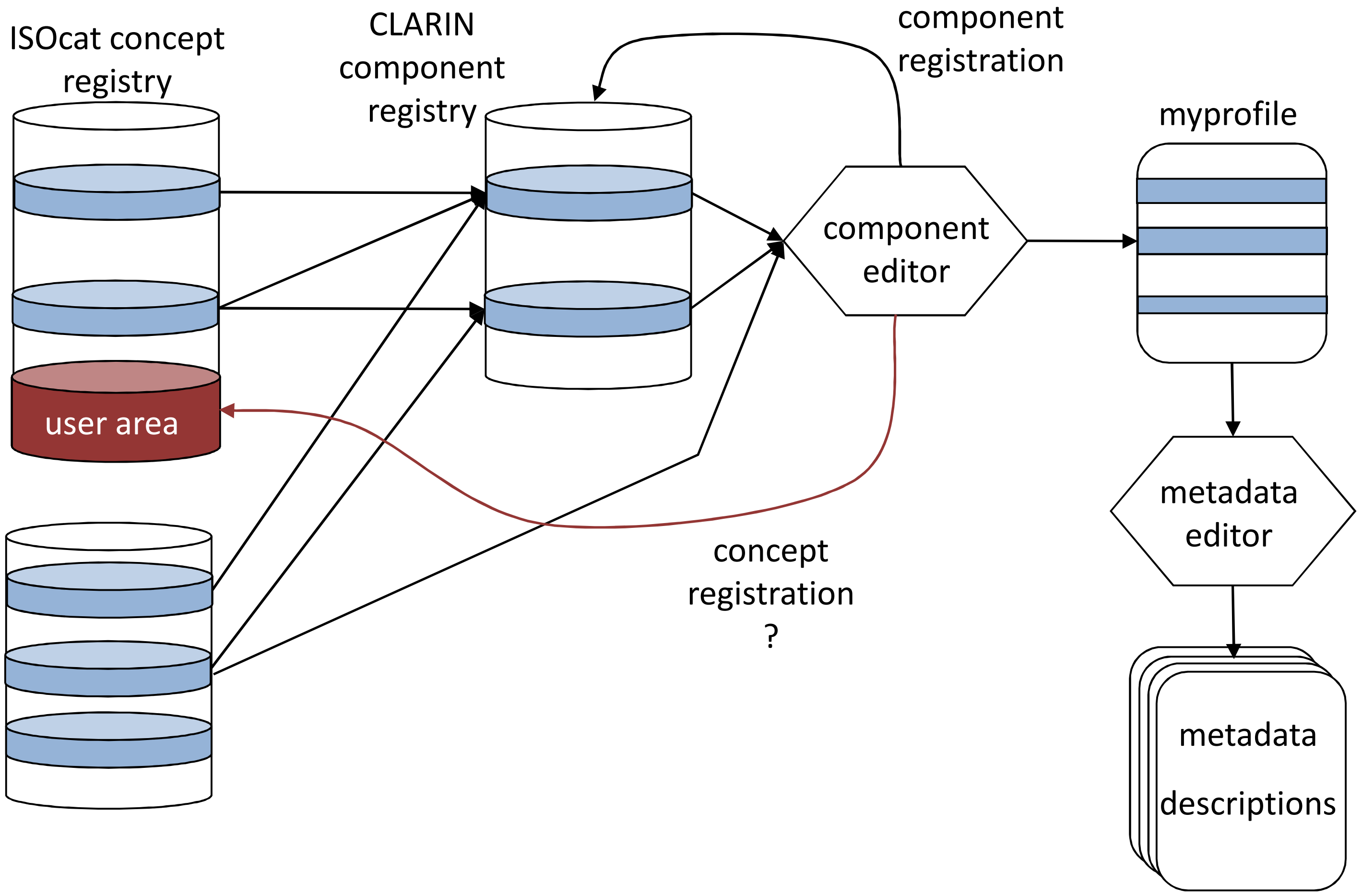
Technological Pillars I

- network of strong centres - 24 serious candidates
  - centres need to meet a number of requirements
    proper repository system, offering standard metadata etc
  - need to participate in quality assessments (data seal - DANS)
- service centre federation
  - allow building virtual collections etc
  - single sign-on, single identity principles
  - establish domain of trust with IDFs
  - intensive discussions with eduGain + TERENA
  - small start-up federation in 09 (DE, FI, NL IDFs)
- persistent identifier service
  - EPIC (European PI Consortium): GWDG/MPG, SARA, CSC, ??
  - based on Handle System
  - only robust, performant registration and resolving system (or?)
  - speak about millions of PIDs (semantic weaving)

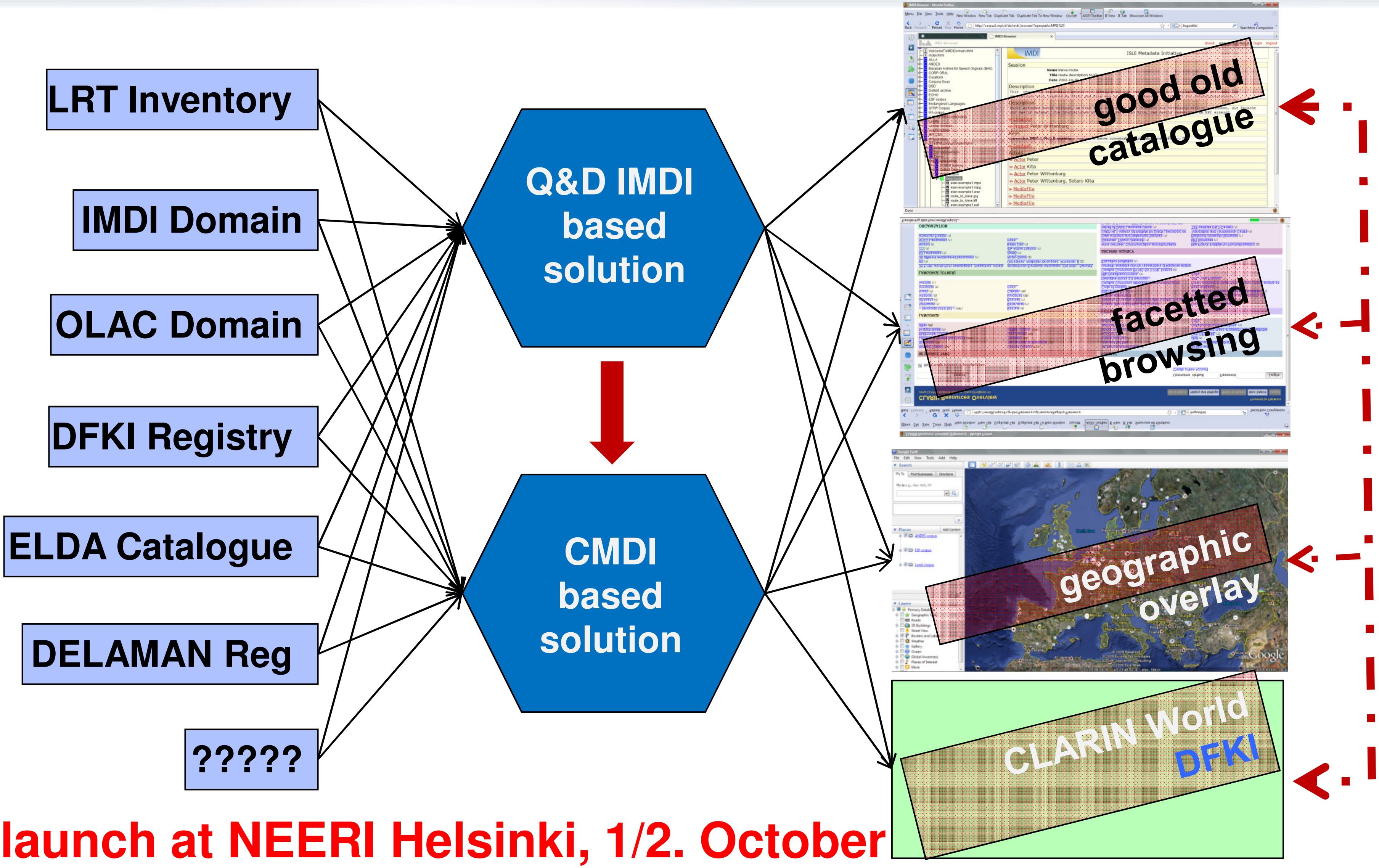
Technological Pillars II

- joint metadata domain based on long experience in the field
  - core principles:
    standardize elements, allow many schemas, use PIDs
  - element and vocabulary registration in ISOcat (ISO 12620)
  - components and profiles to be registered for re-usage
  - harvesting via OAI-PMH
  - five tracks of activities
    - specification, translation of data categories
    - building prototypical components and profiles
    - building component based infrastructure
    - do harvesting and harmonization already now
    - build Virtual Language Observatory (VLO)

CMDI component framework



VLO



launch at NEERI Helsinki, 1/2. October

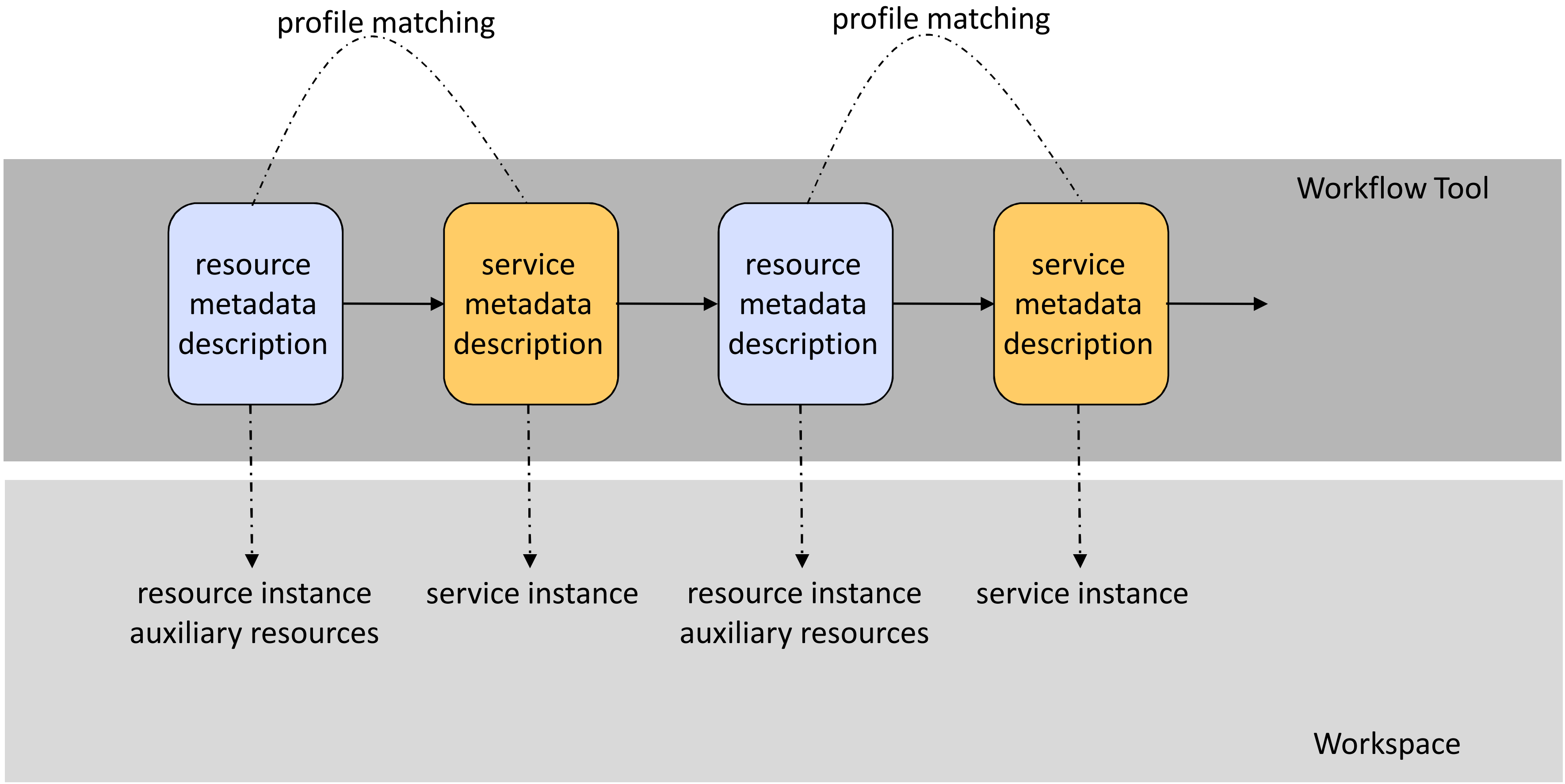
Technological Pillars III

- interoperable domain of LRT - how?
- goal: allow users to build virtual collections and workflows (chaining of web applications and services)

- big issue: standardization and harmonization (ISO TC37, TEI, W3C, ...)
    - quite some standards on resource models on the way
    - great effort to register domain concepts (ISOcat) as basis for future semantic interoperability

- web services/workflow issues
    - basis given by W3C, OASIS etc
    - development of a standard wrapper and service bus implementation
    - which workflow environment ?
        - need asynchronous operation, humans as part of chains
    - working on concrete examples (->Barcelona team and others)
    - now designing a European demo case

MD in workflow chain



profile matching

profile matching

Workflow Tool

resource metadata description

service metadata description

resource metadata description

service metadata description

resource instance
auxiliary resources

service instance

resource instance
auxiliary resources

service instance

Workspace

Gaps

- workspaces for all kinds of activities of infrastructure users

- infrastructure services such as centres registry (separate for CLARIN?)
  busy to design a landscape together with SARA

- execution spaces (close to grid world - what can you offer?)
  - large computation stuff
    training stochastic machines, running complex parsers on huge text collections, automatic annotation of audio/video films, etc
  - small computation stuff - but by many users
    this will be crucial !!!

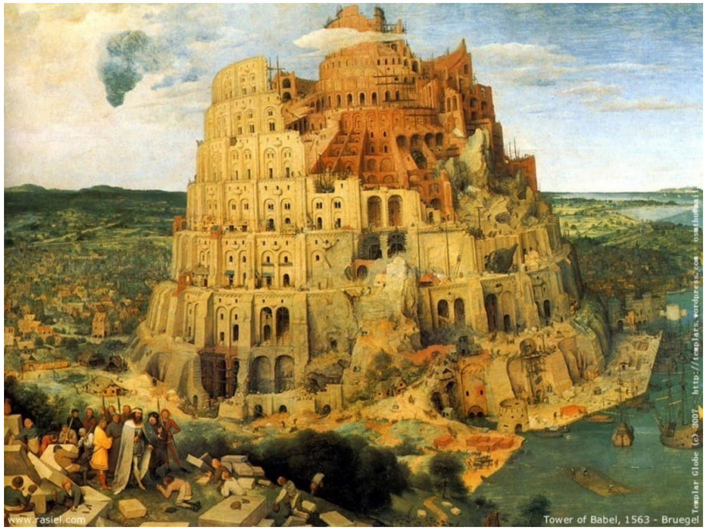
Big Questions

- which infrastructure components are discipline specific?
- which are generic?
- whom can we rely on to give persistent and robust services?

- humanities researchers will only accept if
  - services have high availability and robustness
  - no new burocracy will hamper work (rights issue)
    access patterns in humanities are random!!!
  - they can manage complexity

End

Falls nicht to end in Babylonish scenario nous avons still een beten time om schattingen te improve.


www.rasiel.com
Tower of Babel, 1563 - Bruegel
Templar Globe (c) 2007 - http://templars.wordpress.com -

Thanks for your attention!

www.clarin.eu
www.clarin.eu/VLW
NEERI Conference - Helsinki, 1/2. October 09