

MAX-PLANCK-GESELLSCHAFT

LAMUS: Language Archive Management and Upload System

Alexander König
The Language Archive – Max Planck Institute for Psycholinguistics
Nijmegen, The Netherlands

MPI Archive Data

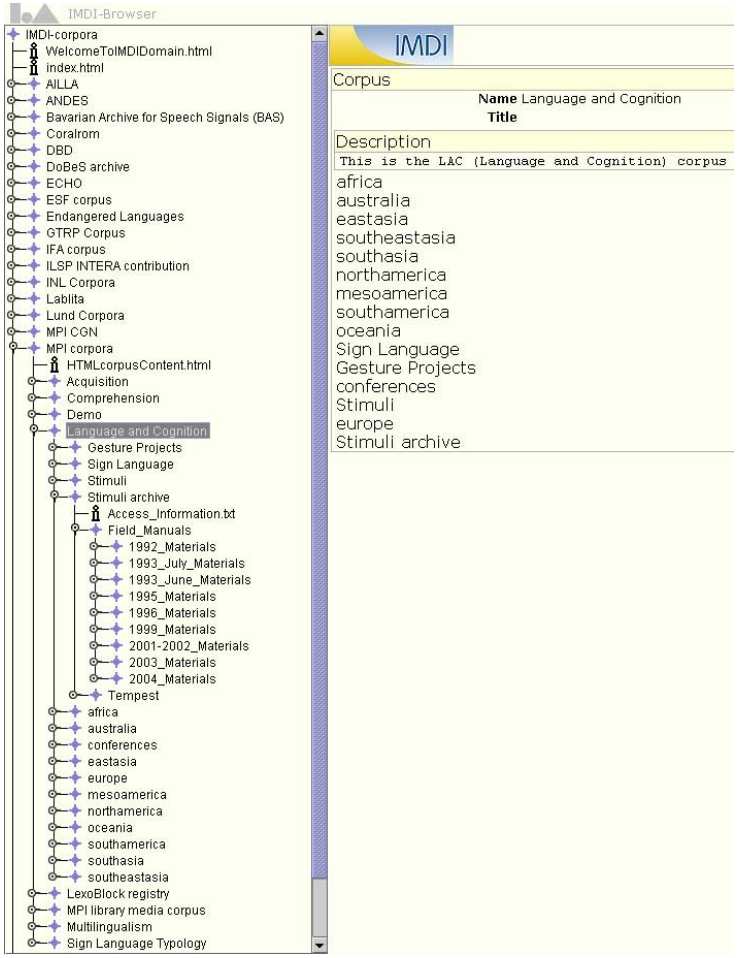
- Large increase in archivable data (currently ~ 55 Terabyte, almost 200.000 sessions, > 500.000 resources).
- MPI archive is open for external donators

How to keep control over this amount of data?

LAMUS

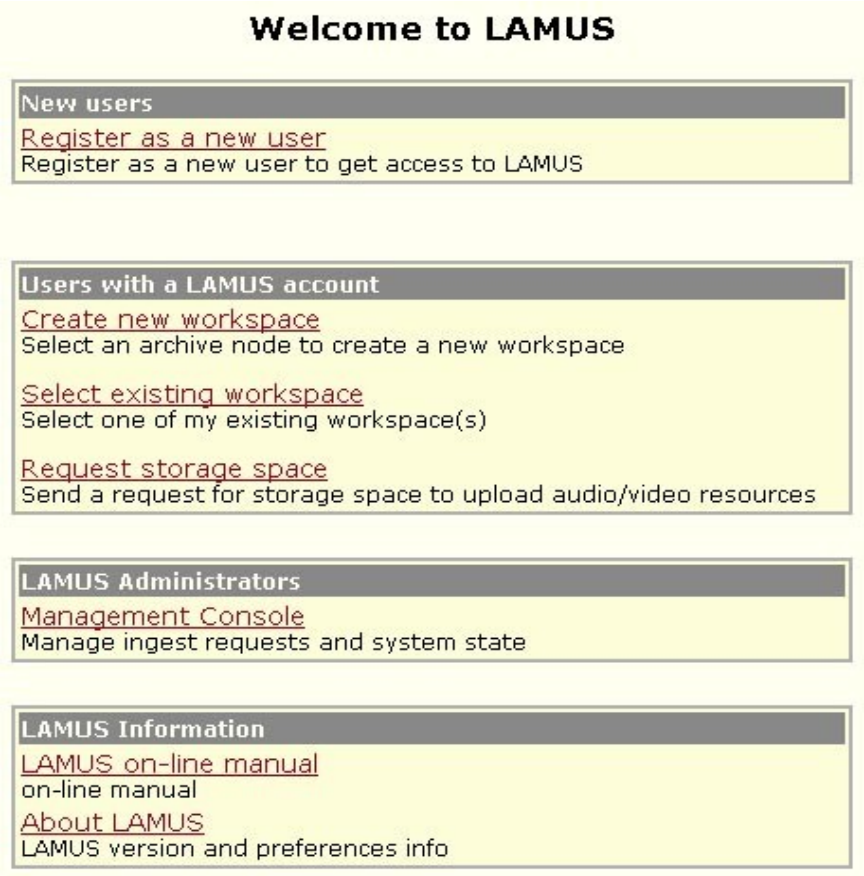
Language Archive Management and Upload System: Researchers can build corpus/ upload resources.

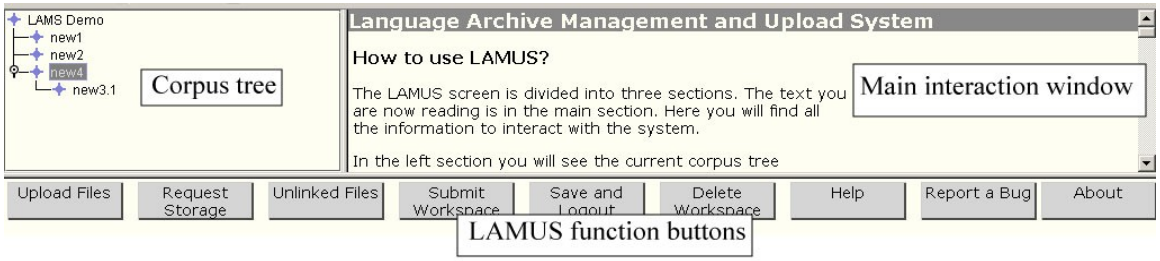
LAMUS checks formats, consistency (gatekeeper)



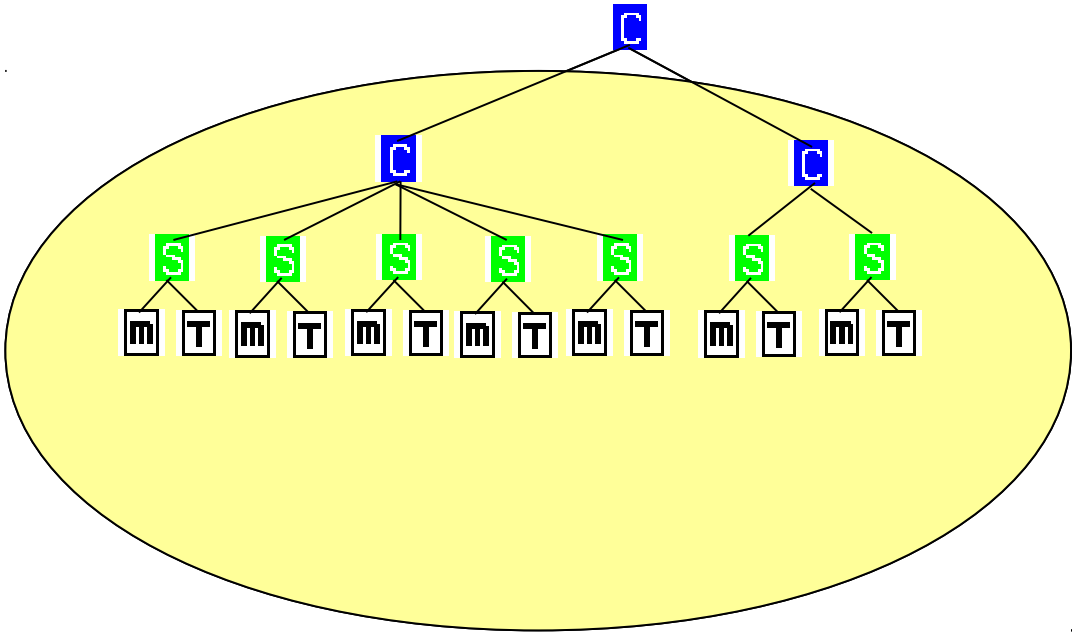
LAMUS features

- Web-based operation.
- Workspace principle (isolated work environment).
- Create and modify the archive's IMDI corpus structure.
- Replace existing metadata and resources in the archive.
- Upload new metadata and resources into the archive.
- Gatekeeper function: not all data formats are accepted; keep archive consistent and future proof.
- Persistent identifiers, Versioning system





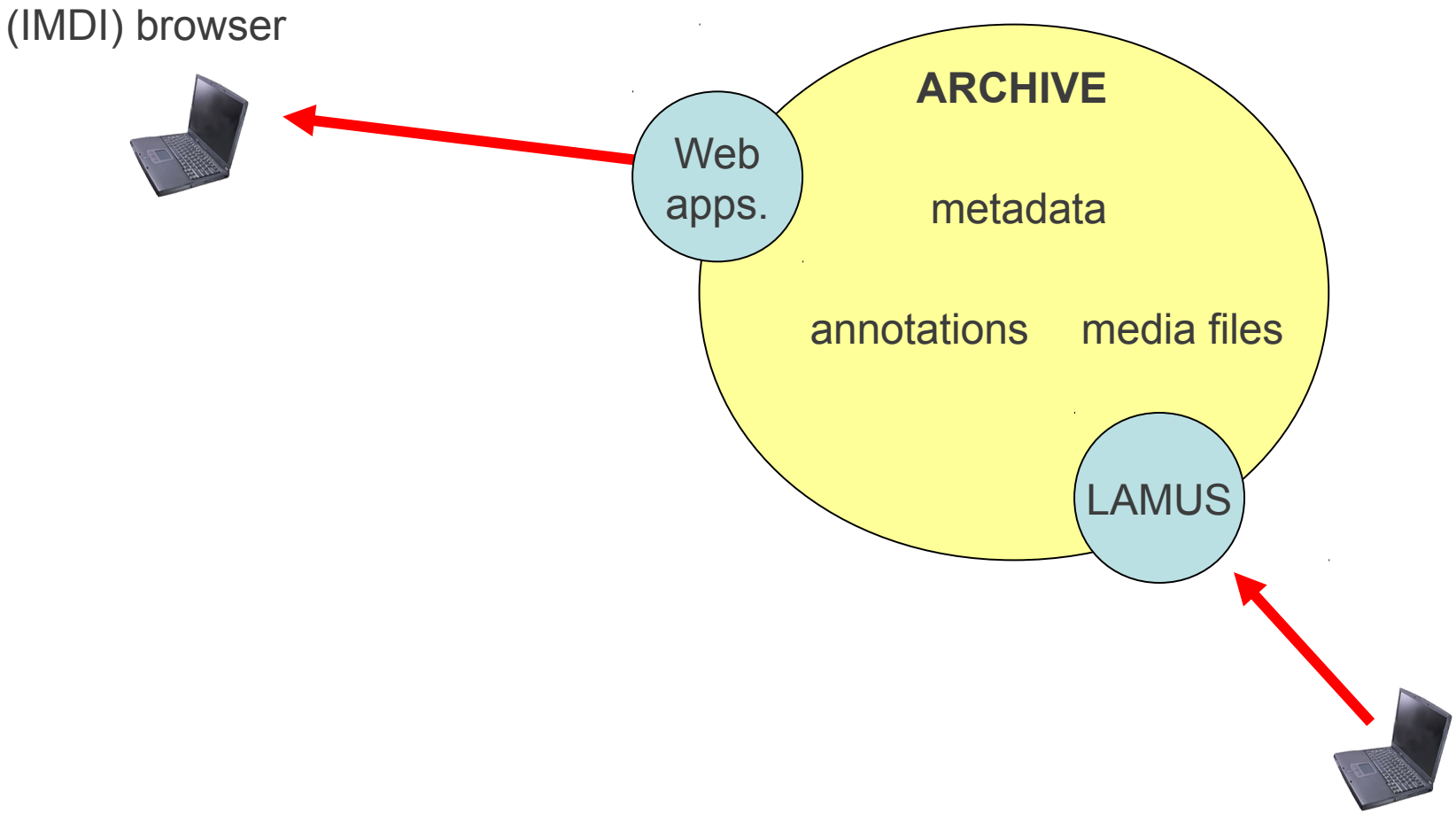
Archive Organization: IMDI domain



- C Corpus node: defines structure, linked session nodes.
- S Session node: detailed metadata description, linked resources.
- m T Resource: text, multimedia.

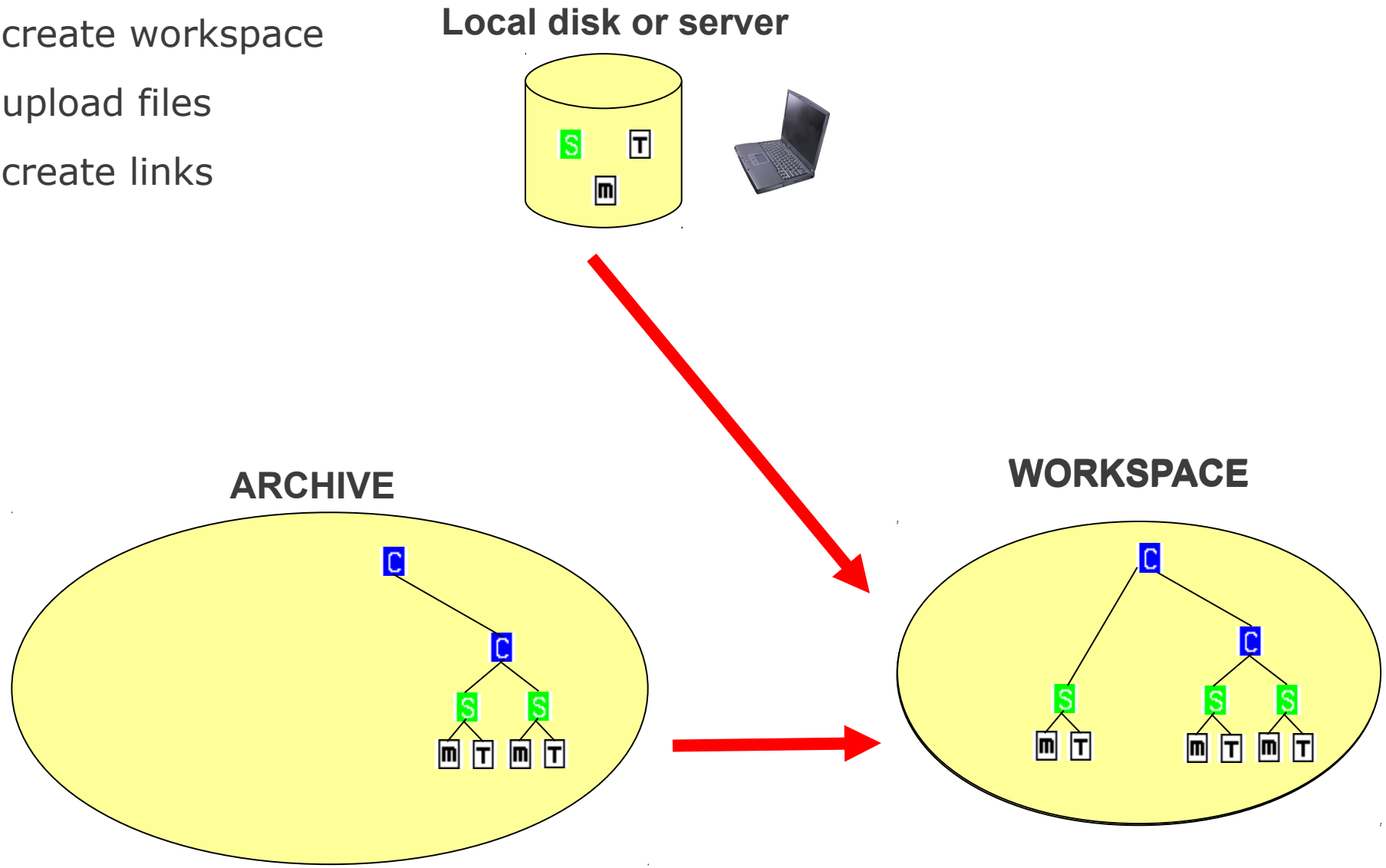
MPI archive organization

Interaction between researcher and archive: browsing/search/visualization



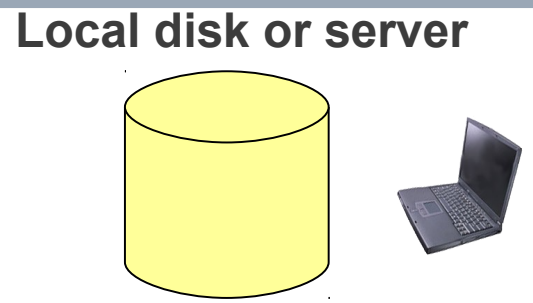
Work Flow Overview

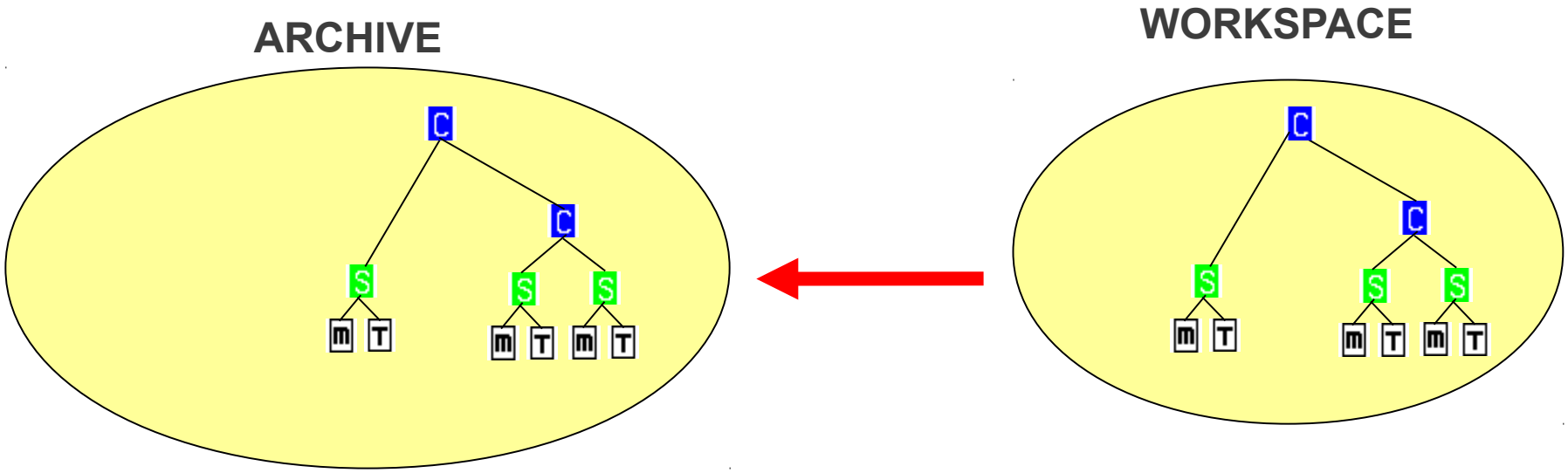
- create workspace
- upload files
- create links



Work Flow Overview

- create workspace
- upload files
- create links
- submit workspace





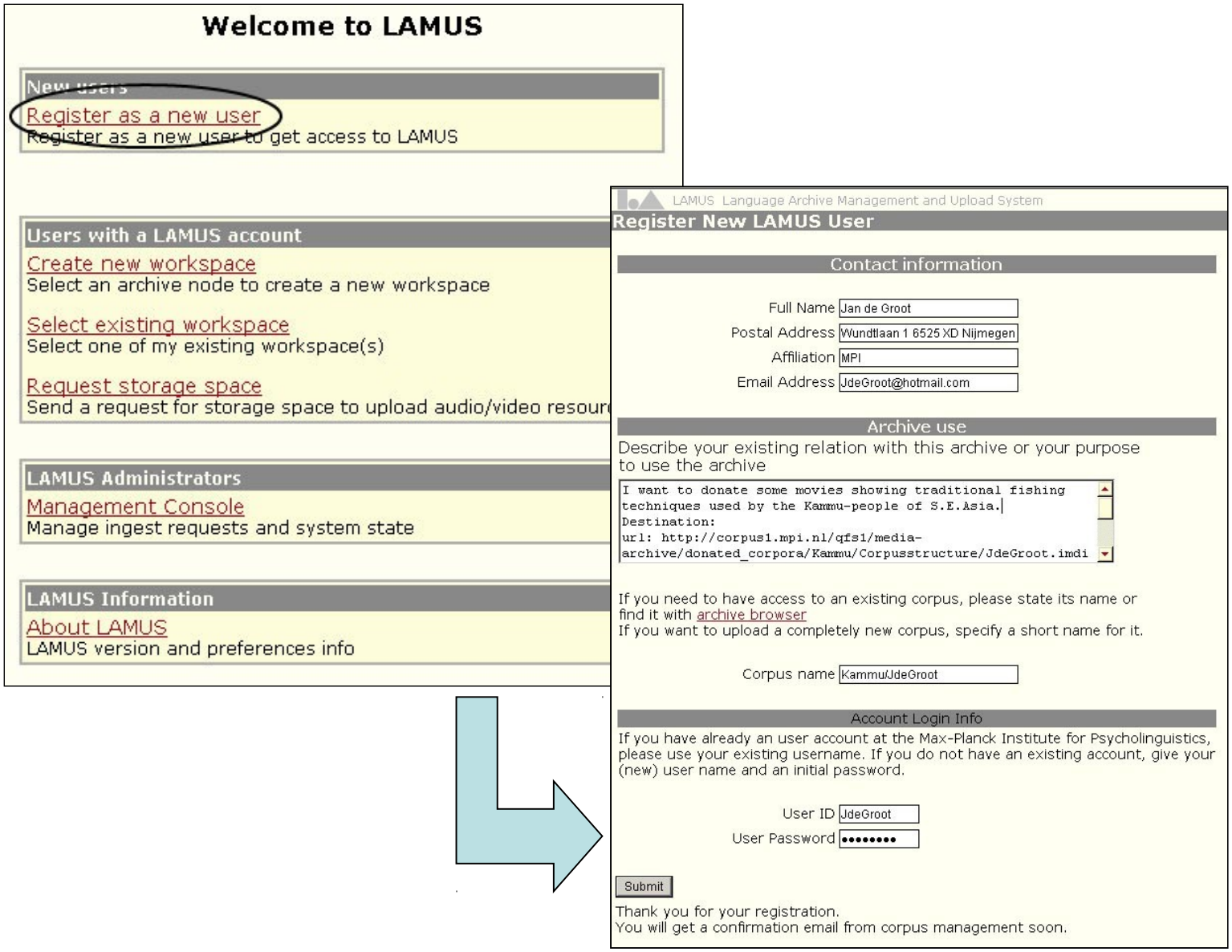
Work Flow Overview

- Open LAMUS in web browser: http://corpus1.mpi.nl/jkc/lamus
- 1st time: ask corpus manager for write rights.
- Select your workspace (as low in the tree hierarchy as possible!)
- Upload metadata, resources, change corpus structure, etc.
- Submit the workspace = apply changes to actual archive.

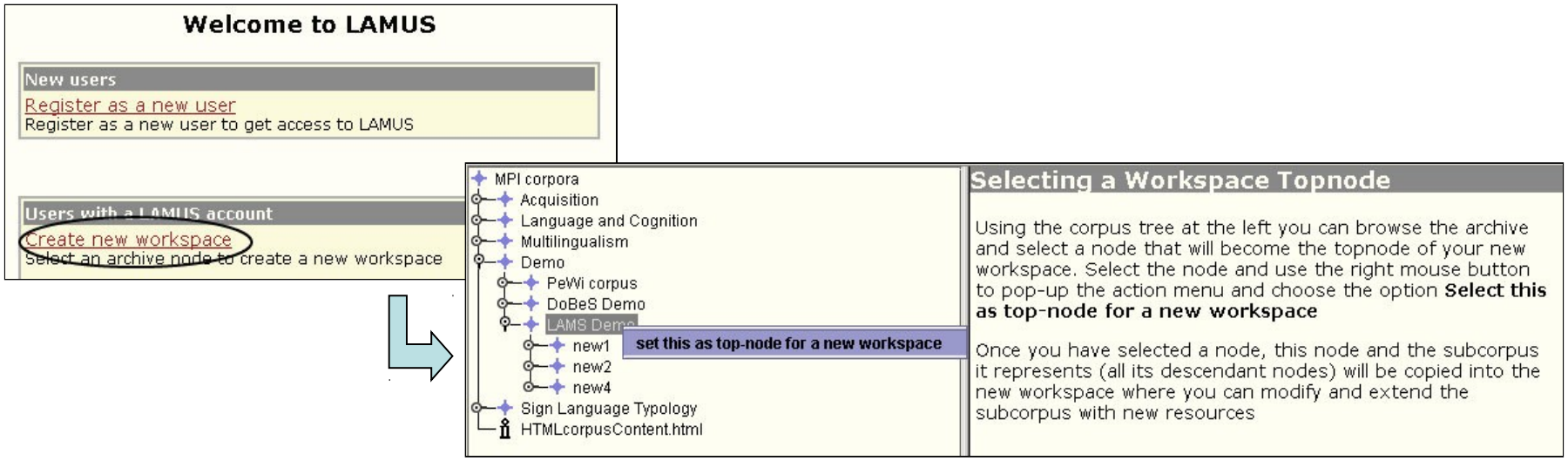
  Optional:

  - Manage access rights to the resources using AMS (metadata is ***always*** accessible for everyone).

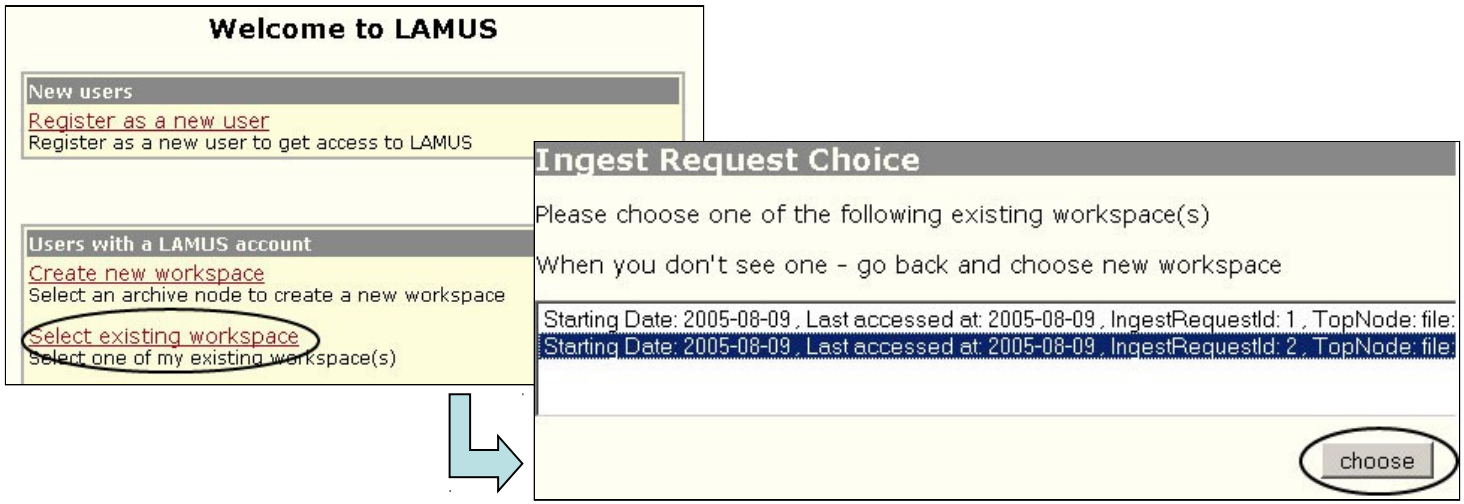
Register as a user



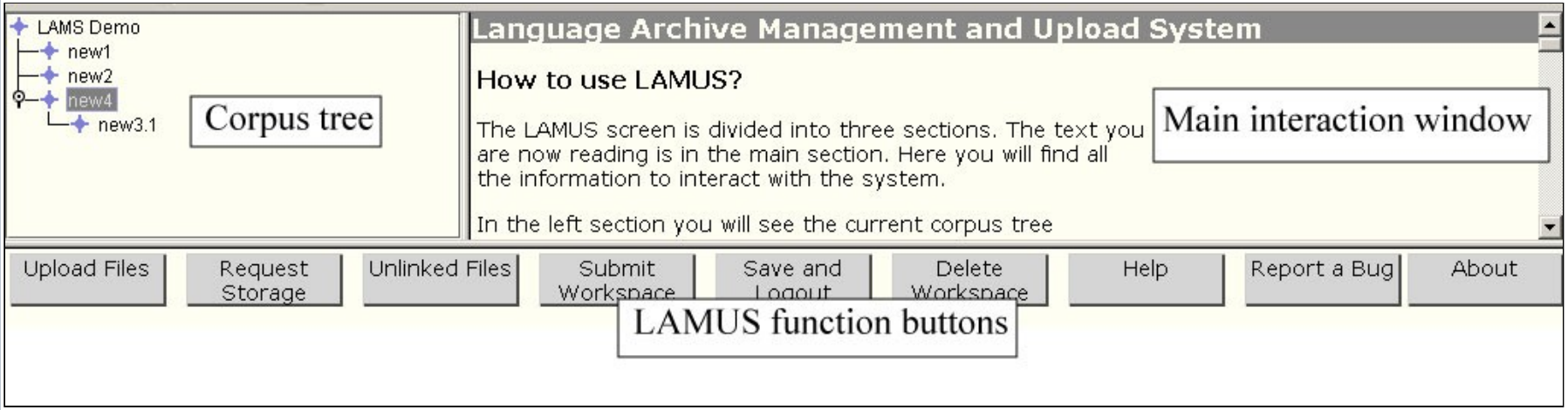
Select a new workspace



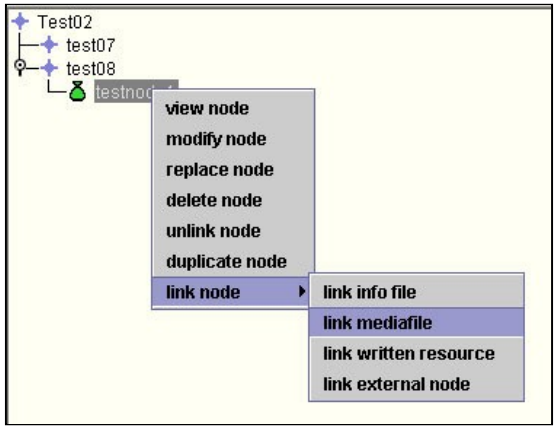
Open existing Workspace



Manage the workspace



Right mouse button on a node: context sensitive tree menu.

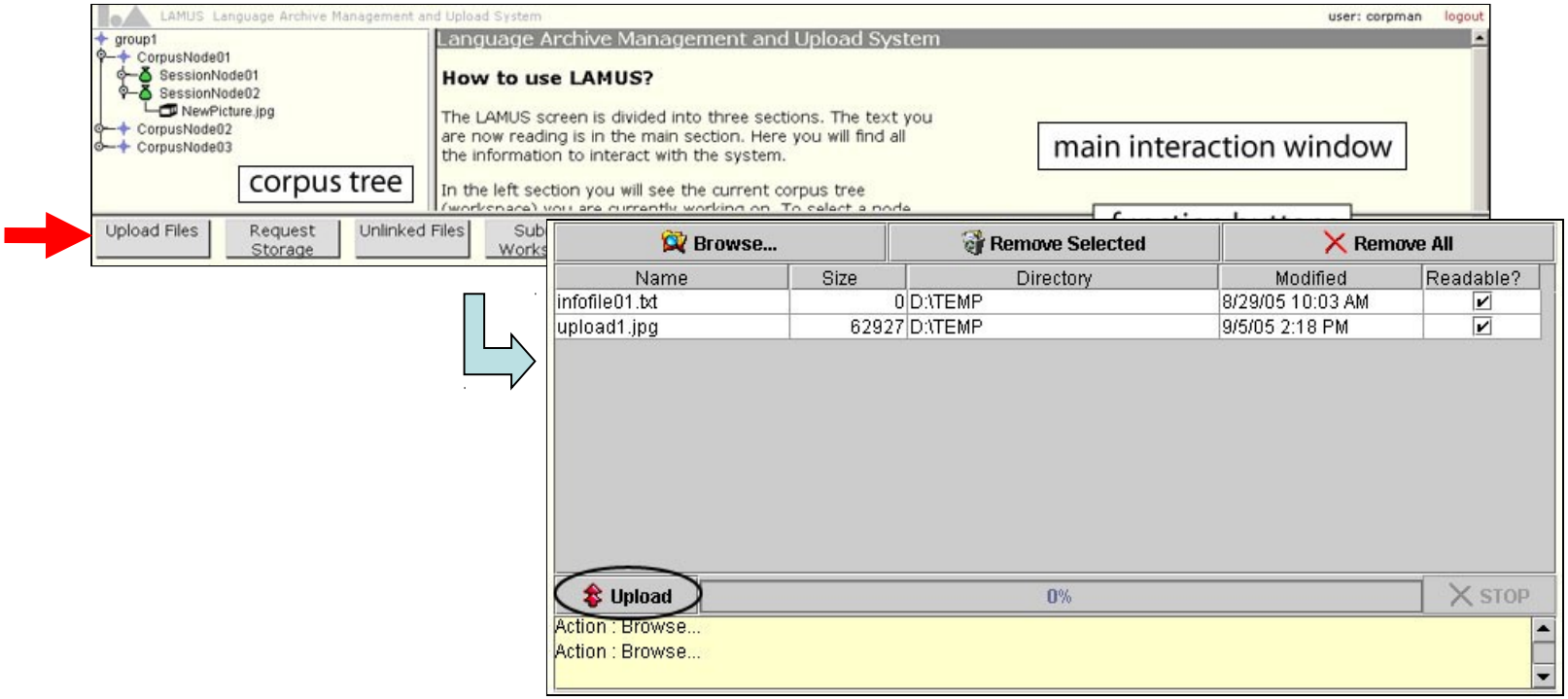


Options like:

- Adding corpus nodes, changing structure
- Renaming
- (Un)linking session nodes, resources

Example: add a new resource

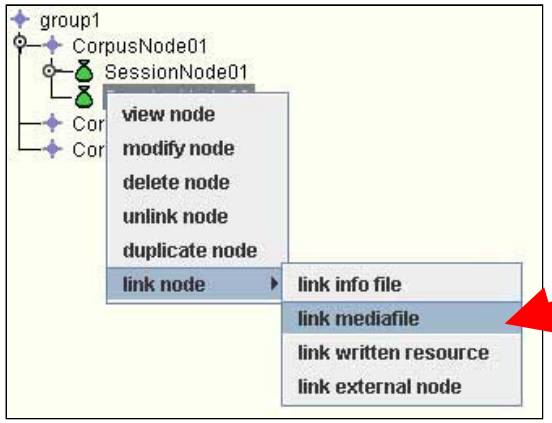
Step 1: upload the local resource to the workspace.



After upload the file types and formats are checked, only accepted ones will be placed in the "Unlinked files" container.

Example: link resource

Step 2: Link the resource to a session node.

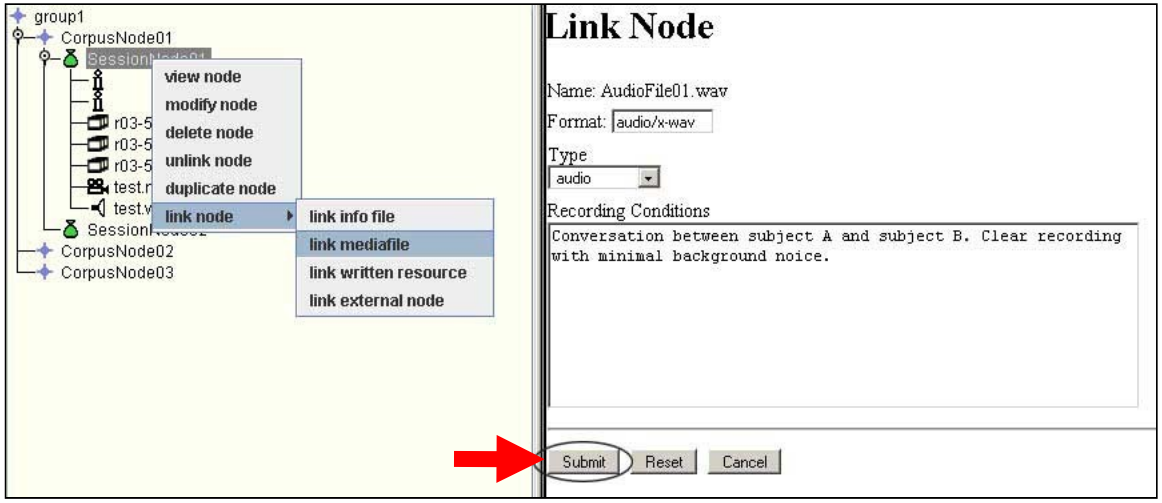


Choose file type.



Example: link resource

Step 3: add additional information and submit



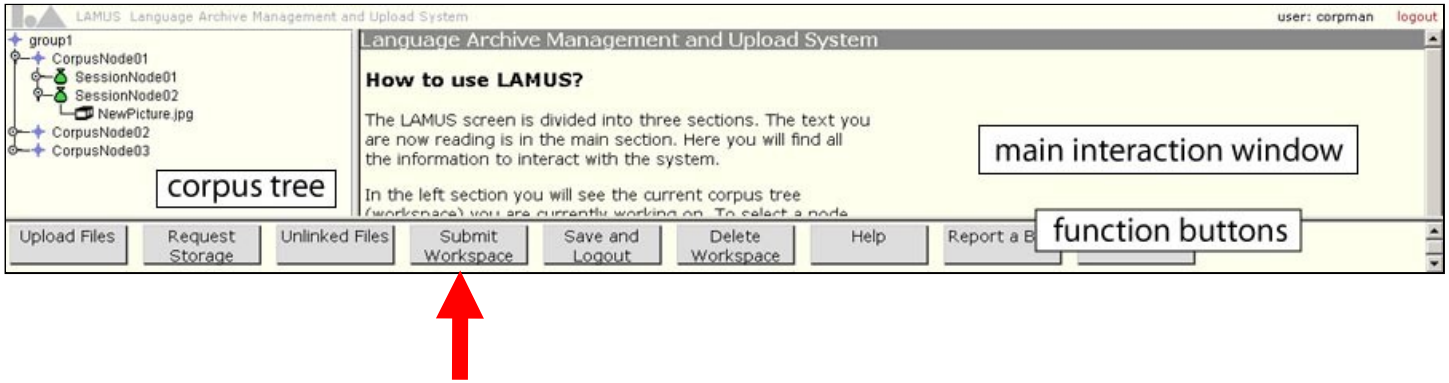
The resource should now be linked to the session in the corpus tree of the workspace.

Example: submit workspace





Step 4: Submit the workspace into the actual archive





Summary



LAMUS makes it possible to:

1. Allow researchers to manage their own corpora.
2. Check for consistency.
3. Update databases and indexes (used by search engines).

Be your own corpus manager!

More information at:

http://tla.mpi.nl/tools/tla-tools/lamus

- Link to LAMUS
- Manuals
- Fora