



DAM-LR: Objectives and State



Peter Wittenburg

MPI for Psycholinguistics



ESF/NSF cyberinfrastructure meeting



- just returned from a Cyberinfrastructure meeting of the ESF/NSF in Paris Monday/Tuesday where I was invited as guest speaker
- it was about data management of IPY and HSS programmes
- James Herbert (Responsible for US HSS Cyberinfrastructure):
“We are looking with great respect to the infrastructure work in the linguistics area in Europe – it’s absolutely outstanding work that you are doing”
- *my answer was: “after your talk I finally know what we are actually doing – what the wider perspective is”*
- *it is very obvious that other HSS disciplines and programmes are behind and that we are setting the trends!!*



some of the major problems



- according to a UNESCO overview about 80% of our recordings about cultures and languages are endangered
- the amount of high quality digital resources about cultures and languages is exploding
- according to the speaker > 80% of our digital material on notebooks and PCs is highly endangered
- according to the speaker ?? % of our digital material is not organized, catalogued - therefore invisible and thus inaccessible (there are no real centers!)
- researchers have to cope with all sorts of boundaries and restrictions (*commercial, institutional, structural, semantic, etc*)



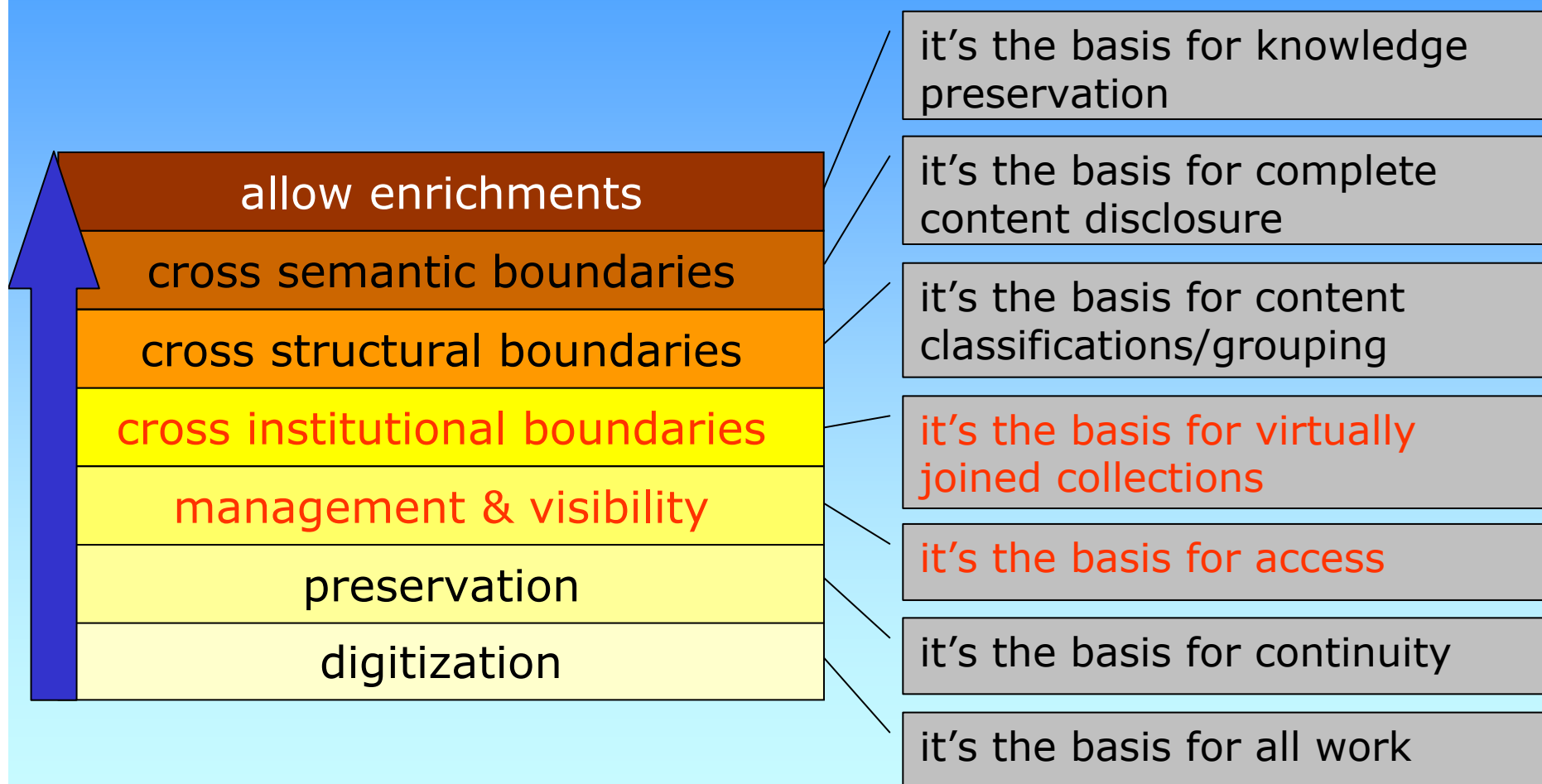
some of the challenges



- large digitization programs
- clear trend to centers storing digital resources (overcome digital trap)
- preserve cultural heritage for future generations
 - not simple, not a technological issue, but organizational + political + luck
 - technological solution not new: continuous migration + copying
 - cost need to be extremely low (10% law for bit-stream preservation)
- manage large amounts of data and make them visible
 - standardized XML schema based open metadata
- allow users to access data across institutional boundaries (Grid)
- overcome structural and semantic interoperability problems when accessing data (Semantic Web)
- allow enrichments according to the **Live Archives** ideas
- this all means support of the Open Access idea

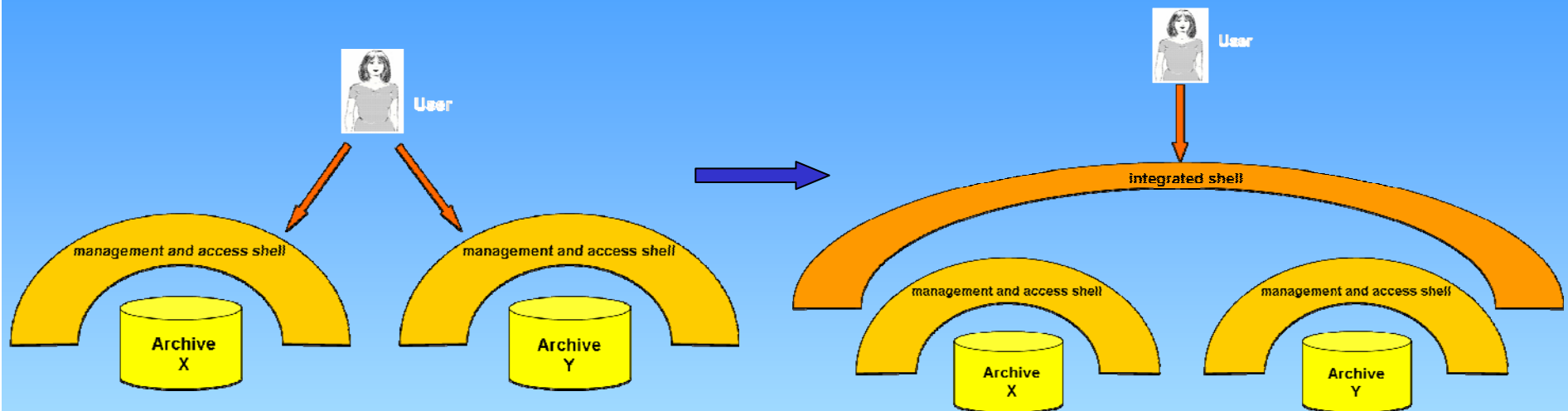


layered system of dependencies





crossing institutional boundaries



- allowing users to create virtual collections by using resources from different language archives/centers
- First priority: establish suitable archives
- Second priority: getting them together
- Third priority: think about legal frameworks
- Fourth priority: think about continuation
- all integration was based on comparatively little money



DAM-LR pillars



- all based on trusted servers, i.e., mutually accepted certificates according to the TERENA TACAR list with certificates signed by PKI system
- joined, but distributed metadata domain with several portals and navigation & searching
- joined, but distributed resolving of unique and persistent resource identifies, i.e., a stable system of referencing
- distributed authentication & authorization system supporting single identity and single sign-on
- moving steps forward towards a federation of archives, i.e., understand the legal and ethical problems
- we tackled all aspects and can demo them, it is not yet stable and fully implemented, completion/optimization required in 07



DAM-LR pillars

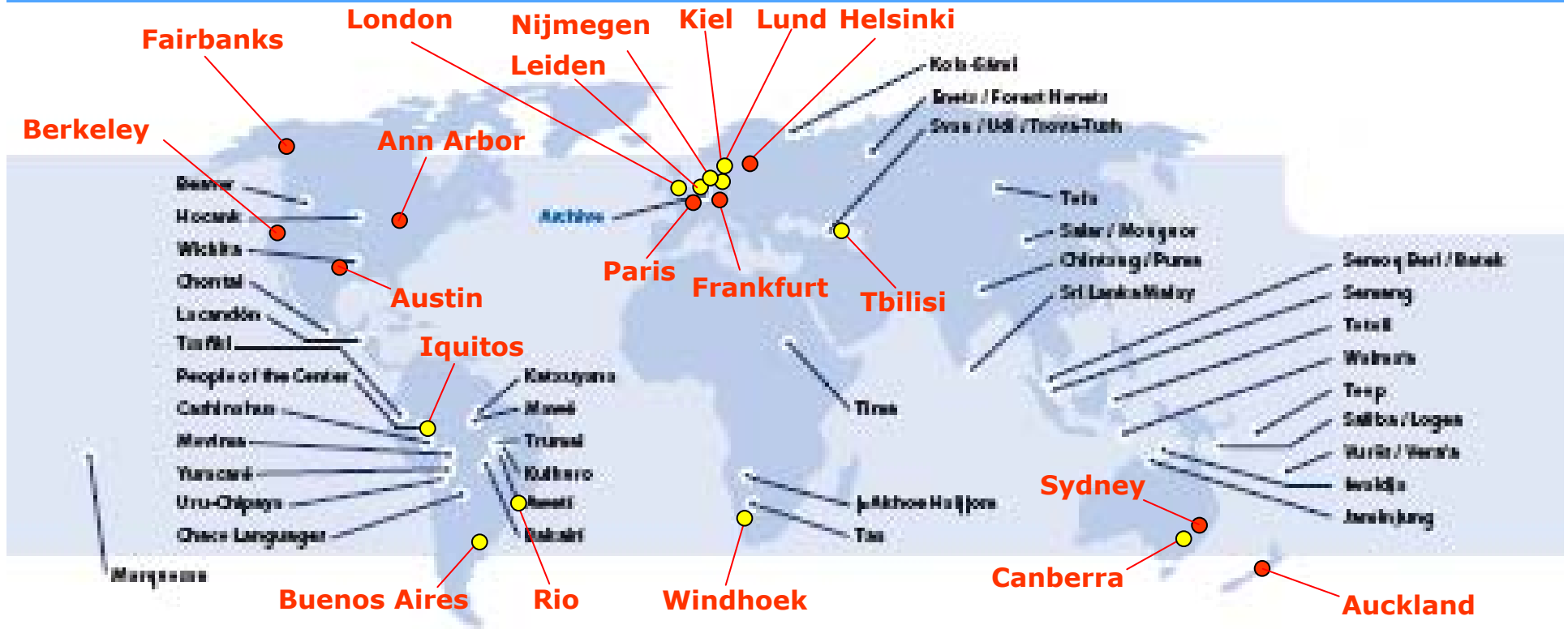


professional archives	all done except SOAS, they had severe problems, but now coming	will continue
trusted servers	all partners requested certificates, partly used in the demo -> Daan	to be fully used
joined MD domain	done, now also including SOAS	will continue
URIDs	“all” installed Handle System	need redundant services
adaptation of local authentication	“all” done	stabilization and adding guests etc
installation and integration of Shibboleth middleware	“all” done and demo -> Daan	needs stabilization and testing
archive federation ideas	did quite some evaluation work	needs a first description of rules

more details during demo session



the DAM-LR Grid at the end of 2007?



- the yellow spots show the DAM-LR group and the LAT archives
- the red spots show our DELAMAN colleagues who follow what we are doing and who have great interest



GIS as the meeting point



- GIS “Language Sites” as an excellent meeting point



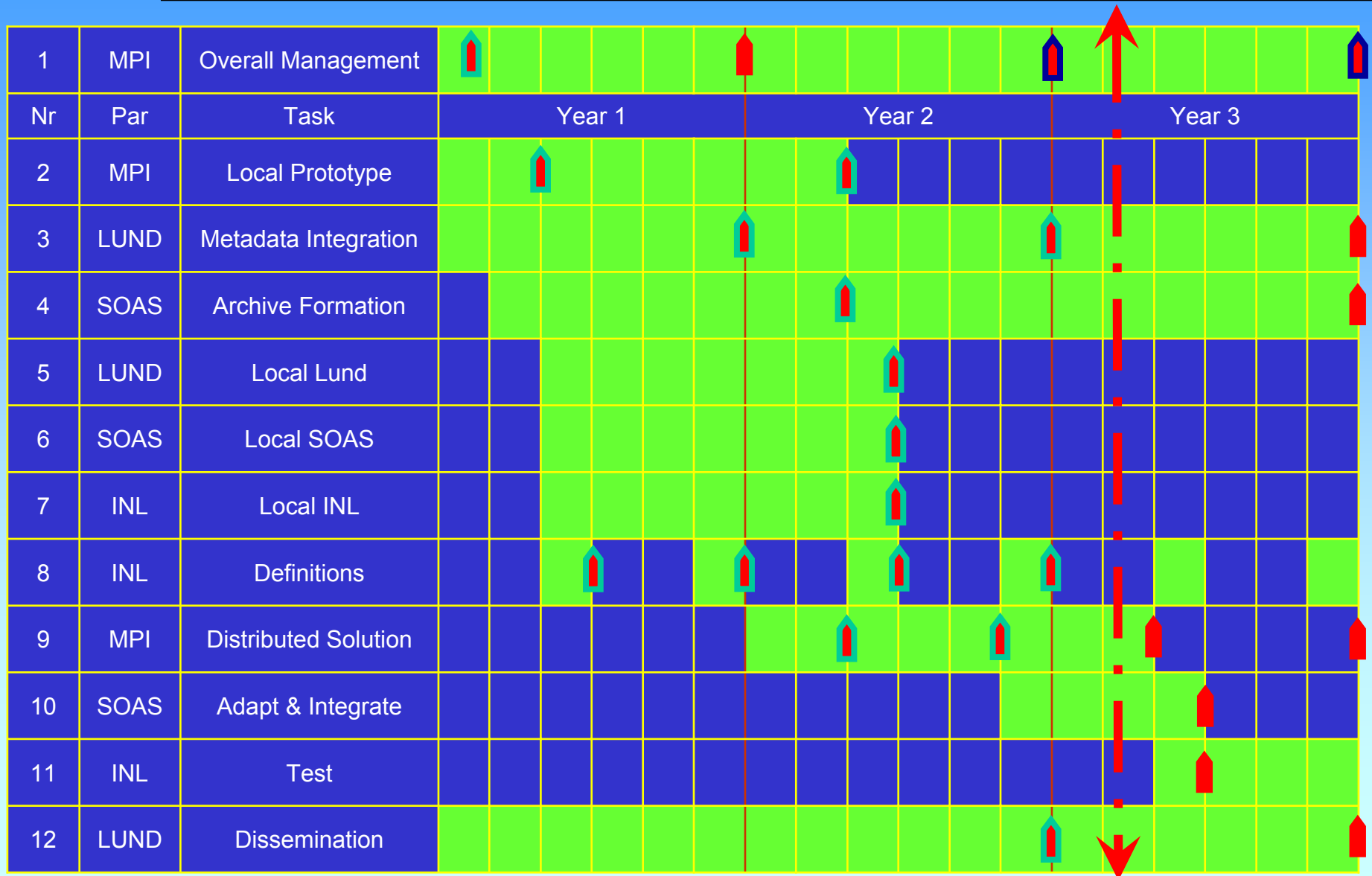
role of partners

	MPI	Lund	SOAS	INL
prototypical archiving system	XX			X
archive formation	X	X	XX	X
keeping metadata in focus	X	XX	X	X
local archive system setup		X	X	X
distributed solution and adapt	XX	X	X	X
test	X	X	X	XX
keep dissemination in focus	X	XX	X	X
keep training in focus	X	XX	X	X
federation discussion	X	X	XX	X

- two original splits could not be made in practice



Time planning & Deliverables





the wider perspective



- DAM-LR is a very small project in terms of partners etc
- but
 - we have an answer to data management – still a gigantic issue
 - we have an answer to user management – also a gigantic issue
 - we start understanding the interplay between centers
 - we wanted to know whether the humanities can setup a Grid system AND YES THEY CAN
(was and is not simple – but we understand now what to do not only technologically!)
 - we are at the basis of something which in 50 or 100 years will be called the digital revolution!
 - we are a small Lego block in advancing the field



2006 was very intensive for all of us



- DAM-LR partners did a lot in dissemination & evangelization
- several workshops (14) were organized by DAM-LR members with DAM-LR issues as topics!!
- several training courses were given

- it was an extremely intensive year
- it was a year with much outreach also internationally

- all money invested so far is well spent, since we are creating sustainable infrastructures

a hell of work for all of us – so thanks from my side