

THE AUSTRIAN LIBRARY NETWORK AND SERVICE LTD.

The logo for obv sg features the text "obv sg" in a dark grey, sans-serif font. To the left of the text, a series of thin orange lines radiate from a single point, fanning out towards the left edge of the slide.

obv sg

PRIMO IN AUSTRIA

MAXIMIZING FUNCTIONALITY AND MINIMIZING INDIVIDUAL EFFORTS BY
RESOURCES SHARING WITHIN A HOMOGENOUS CONSORTIUM

WOLFGANG HAMEDINGER, VICTOR BABITCHEV

THE 5TH IGELU 2010 CONFERENCE 30.08 – 01.09.2010, GHENT BELGIUM

AGENDA

- INITIAL SITUATION
- TOWARDS A SOLUTION
- STATUS
- IMPLEMENTATION
 - objectives / issues & pitfalls / solutions / demo
- EXPERIENCES and TO DO
- EFFORTS (RESOURCES) and COSTS
- NEXT STEPS
- CONCLUSION

Initial situation 2007/2008

Lots of relevant electronic materials

- Electronic journals/articles
- E-books
- Tables of content and other enrichment information
- Theses
- Genuine scientific production of the institutions

Administration of material via different and not smoothly interacting systems

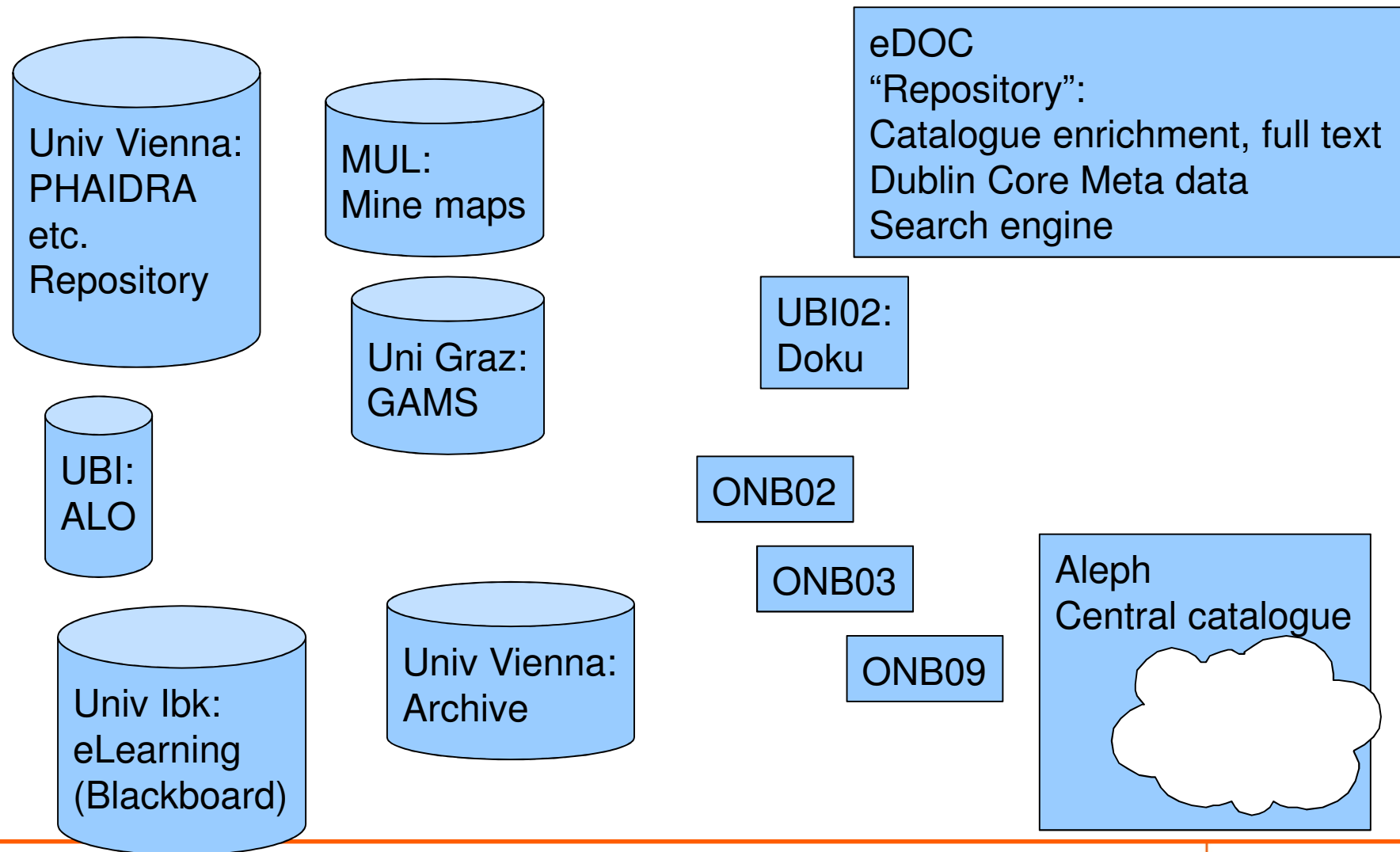
Initial situation 2007/2008

Insufficient meta data information and indexing

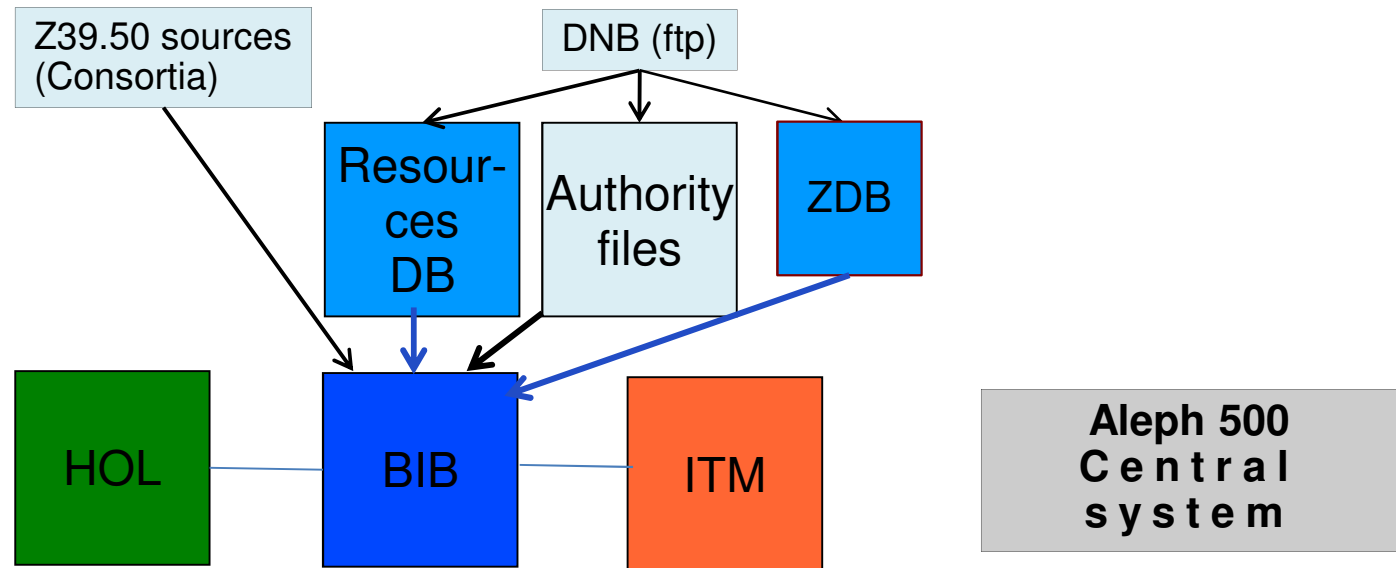
- Only parts of the existing information are catalogued and indexed
- Enrichment information for information available only after finding the record
- Different search strategies and systems necessary
- Unsatisfying usage of expensive material due to lack of search facilities

Not only students but also scientists are more and more using Google and Co for looking up research material

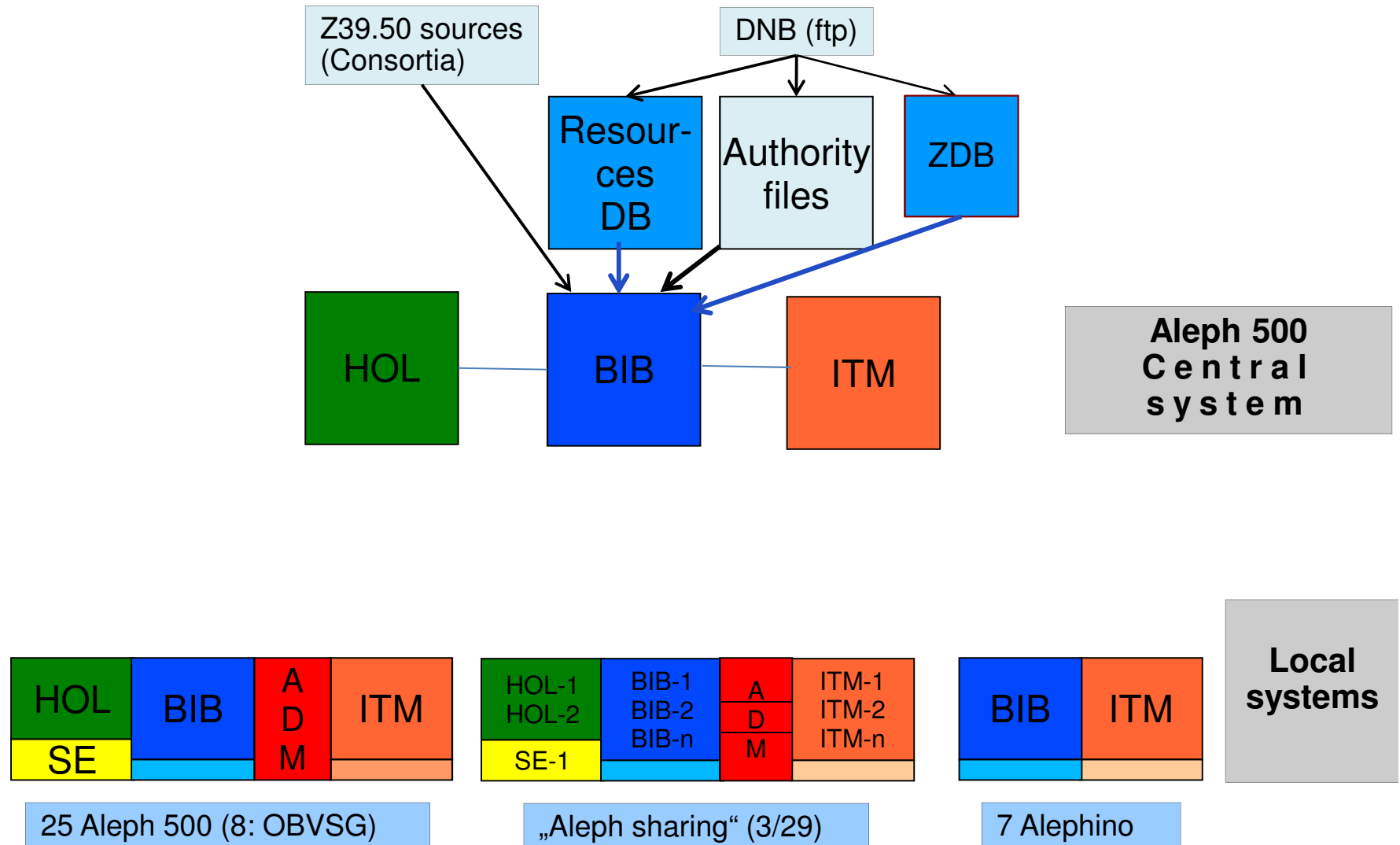
Data sources: a sample of materials within the consortium



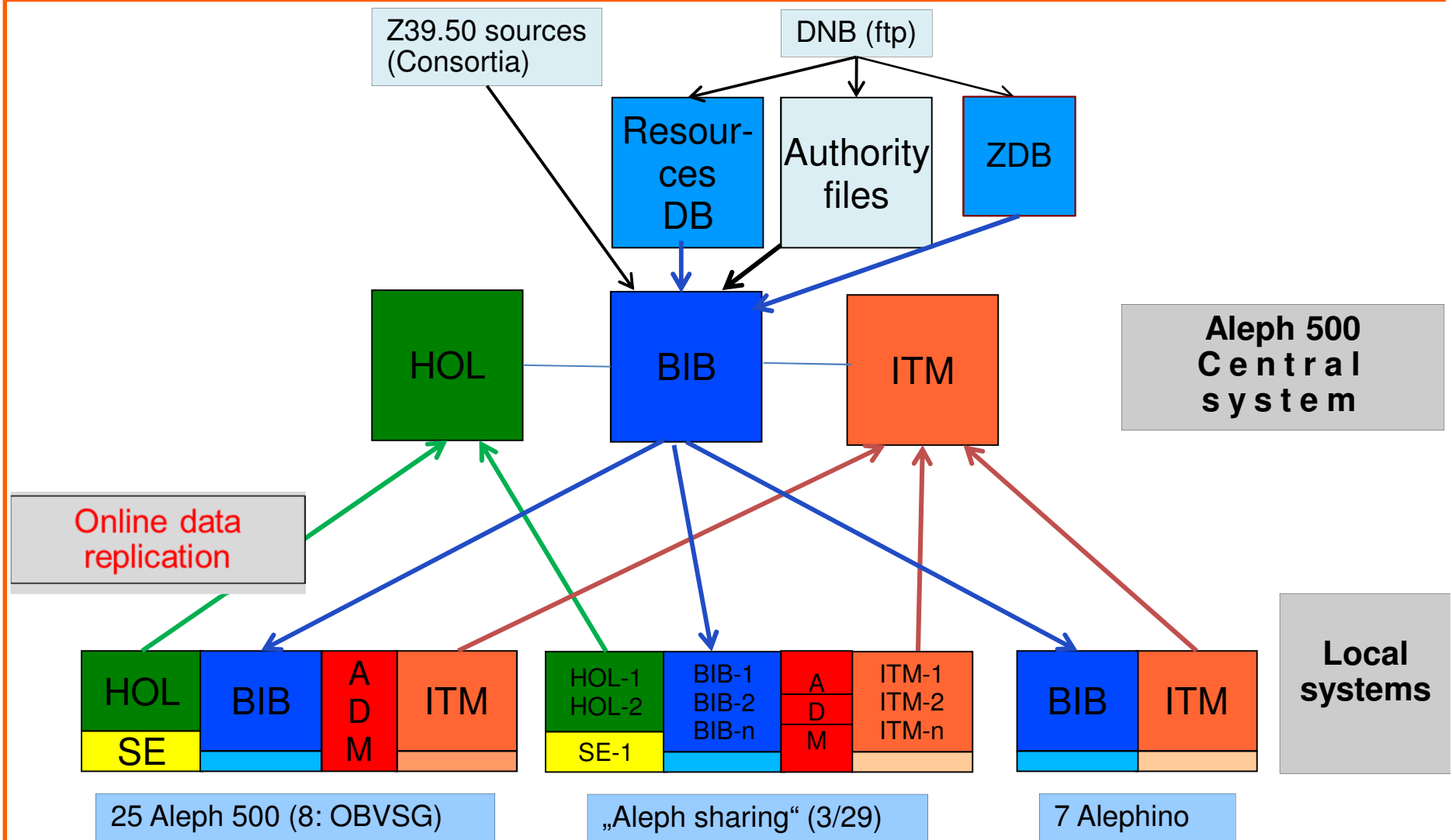
Consortium data source and data flow [1]



Consortium data source and data flow [2]



Consortium data source and data flow [3]



Remarks on the architecture

Almost homogeneous structure

- Started with pure Aleph-systems
- in the meantime also Alephino-systems included

Dispatch of features over one central and many local systems with the general guideline

„As much central as necessary, but as much local as possible“

this provides

- **Flexibility**
- **High degree of synergy**

Economic constraints

Central database without duplicates as policy – not a union catalogue for lucky historical reasons

- Specialised on cataloguing using all possibilities like authority data, resource databases, Z-sources
- Catalogue enrichment
- Other services of general use and interest

Bidirectional data replication between central system and local system

- Every user benefits from data improvement automatically

Economic constraints

Changes in law – no more possibilities of direct central funding by the ministry for libraries

- Newly proposed projects are in competition with other university duties
- Necessity of self organization within the consortium
- So we have to use all possible synergies of our consortia architecture

RECOMMENDATIONS

Recommendations of working group to general assembly

1. Use of search engine technology in conformance with general user behaviour
2. “Completeness”: Cover all relevant materials of an institution and use it from a single search point (“One-stop-shop”) [means not only the library!]
3. “Accuracy”: cover only material, which can be immediately delivered by the institution
4. Include article data with abstracts and catalogue enrichment
5. Use all assets of the existing consortium architecture
 - a. Uniform processing mechanism for used data
 - b. As much central implementation as possible

Recommendations of working group to general assembly

6. Use a model for harvesting data from distributed central/local repositories
7. Create local views with institutional "Corporate Identity", which respect available material and licence situation
8. Use full text indexing for very special and manageable material of the institutions [like theses and own research papers]
9. Fund the necessary resources for implementation and permanent operation

September 3rd, 2008

TOWARDS A SOLUTION

“Its the economy, stupid”

Facts:

- Main source of data is the central database, because we use shared cataloguing
- Holdings information resides in the local library systems and in the linking resolver databases
- All data enrichment information is linked to central bibliographic records

“Its the economy, stupid”

Facts:

- Due to the architecture of the consortium we can pull the local holdings information to the central site automatically nearly for free (it needs little bit additional bandwidth and hardware)
- A replacement for the OPAC only will always be too expensive
- The search interfaces have to be flexible enough for our local institutions

“Its the economy, stupid”

Concept:

- Concentrate all efforts on a uniform centralized data preparation method with a big common and homogeneous index pool
- Use all assets of the existing architecture, infrastructure and cooperation
- Establish a centrally implemented and operated installation of this index base

“Its the economy, stupid”

Concept:

- Use appropriate slices out of this pool for single institutions
- Find a multi tenant solution, which gives every institution the responsibility of tailoring their search interface to its needs within the given technical restrictions of a common solution
- Create a central service with standard features, which can be easily reused by new participants

“Its the reality, sometimes also stupid”

Going Primo, Part 1

- Intense discussions started from mid 2007
- Contacts with early adopters at IGeLU conference
- Some serious contacts between potential Austrian customers and Ex Libris in June 2008
- Principle recommendations of the working group
- Study visit at Copenhagen

Primo Version 2 lacks multi tenant capabilities, generates data redundancies and has only a rough concept of full text indexing with many limitations

“Its the reality, sometimes also stupid”

Going Primo, Part 2

- Developer workshop with Ex Libris in Jerusalem, begin of December 2008
 - In depth discussion of situation, needs and direction of development
 - Concept of step by step solution
 - Very intense and fruitful meeting
- Delegates recommended to go with Primo in spite of the current limitations, if development of solution path and appropriate time frame becomes part of the contract

“Its the reality, sometimes also stupid”

Going Primo, Part 2

- Sign of main contract on December 15th, 2008
- Technical tests in January 2009
- Implementation starts with Kick-off meeting on February 26th, 2009

STATUS

Status: Highlights

The implementation of the system lasted a little bit longer than 3 months – about 1 year

We achieved important functionalities

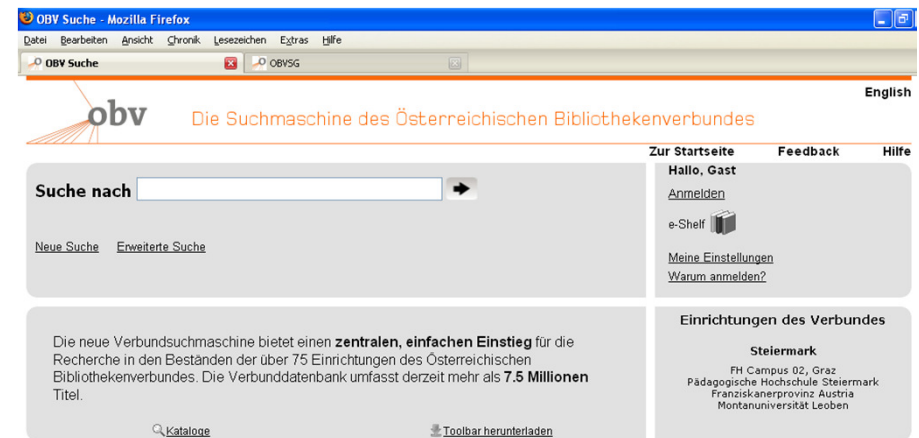
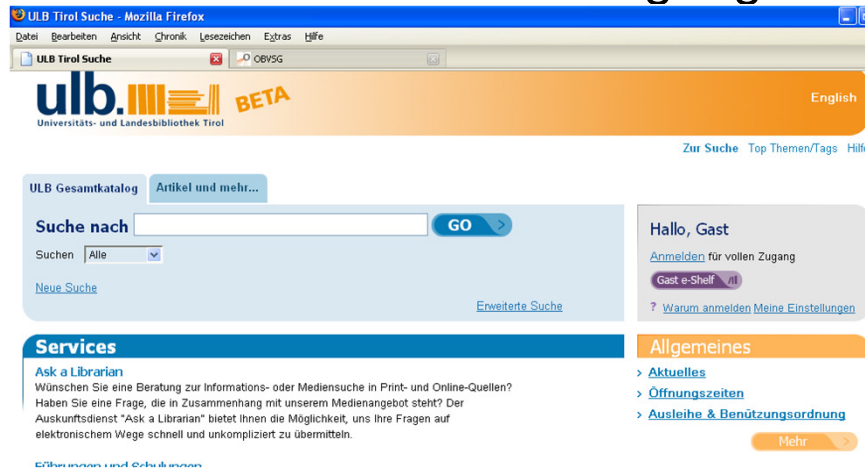
- Consortium Primo with one installation for many participants
- Full text indexing
- Development of standard procedures, work frames and templates for reuse by new Primo users

Implementation phase ended mid 2010 – we are in the full production now

... and it is a rather complex Primo installation ...

Status: What have we achieved?

- Implementation of technical and organisational infrastructure
- 4 Primo views in production
- Full integration of eDOC data
- Integration of MetaLib / SFX
- Enrichment by IVSCAN articles
- Authentication via Shibboleth
- Web 2.0 functionalities + gadgets



Status: Statistics

Status: May 2010

	Consortium BIB	Articles IVSCAN	eDOC	Local BIB	SFX	Meta-Lib	Sum
UBI	1.481.354	272.411	132.985	226.219	42.254	75	2.022.373
UBW	3.000.875	455.131	183.606		62.020	401	3.518.427
OBV	7.619.183	-	301.965		-	-	7.619.183
Total BIB	12.101.412						

IVSCAN: Title information without holdings, transported from consortium to local Primo view (about 650.000 articles!)

eDOC: Enrichment data from eDOC (about 420.000 objects)

Local BIB: Local data (without consortium AC-ID) from UBI01

Status: Next participants

Institution	Production	Version
University of Veterinary Medicine Vienna	few weeks ago	2
Migration to 3		
Vienna University of Economics	2010 Q. 3 - 4	3
University of Graz	2010 Q. 4	3
University of applied Sciences Vienna	2011 Q. 1	3
Austrian National Library	2011 Q. 1	3

PITFALLS DURING THE IMPLEMENTATION

Pitfalls during implementation

How the implementation was running in Austria?

Initial expectations

- sure we will confront some obstacles
- but we will overcome them in any way
- and will come to finish well, perhaps with a small delay



Pitfalls during implementation

Looking back at our most “famous obstacles”

- Project management
 - Lack of communication between partners (initial phase)
 - Steering Committee was set up rather late
 - ExI reaction time on bug fixes was not always satisfactory
 - Frustration, especially when critical issues delayed.....



Pitfalls during implementation

- System administration
 - Analysis of errors and log files was challenging
 - We missed badly a good training and more complete documentation
- Authentication (PDS/Shibboleth)
 - a lot of time was lost – documentation incomplete
 - we missed expertise from Ex Libris



Pitfalls during implementation

- Full text indexing – one of the most critical functionalities
 - We suffered from bugs and undocumented implementation specialties
 - Intense analysis of possible solutions by OBVSG and joint development with ExI of approaches to our problems
- happily that we could work direct with core ExI developers ☺



Pitfalls during implementation

Pitfalls = **higher** project costs!

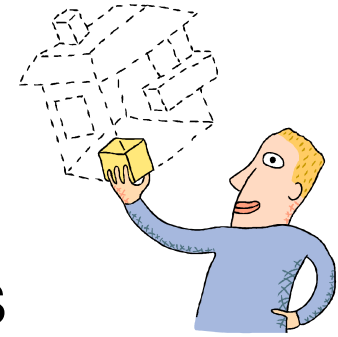


IMPLEMENTATION

Objectives, Issues and Solutions

Objectives 2009 - Base Primo Implementation

- Integrate Central data
- Implement standard Primo functionalities
- Implement solutions and workflows enabling easy integration of new institutions
- High level of automation for all operations
- Operational stability and high availability



Switch Primo Online for the first three Institutions!

Building Base Primo Implementation



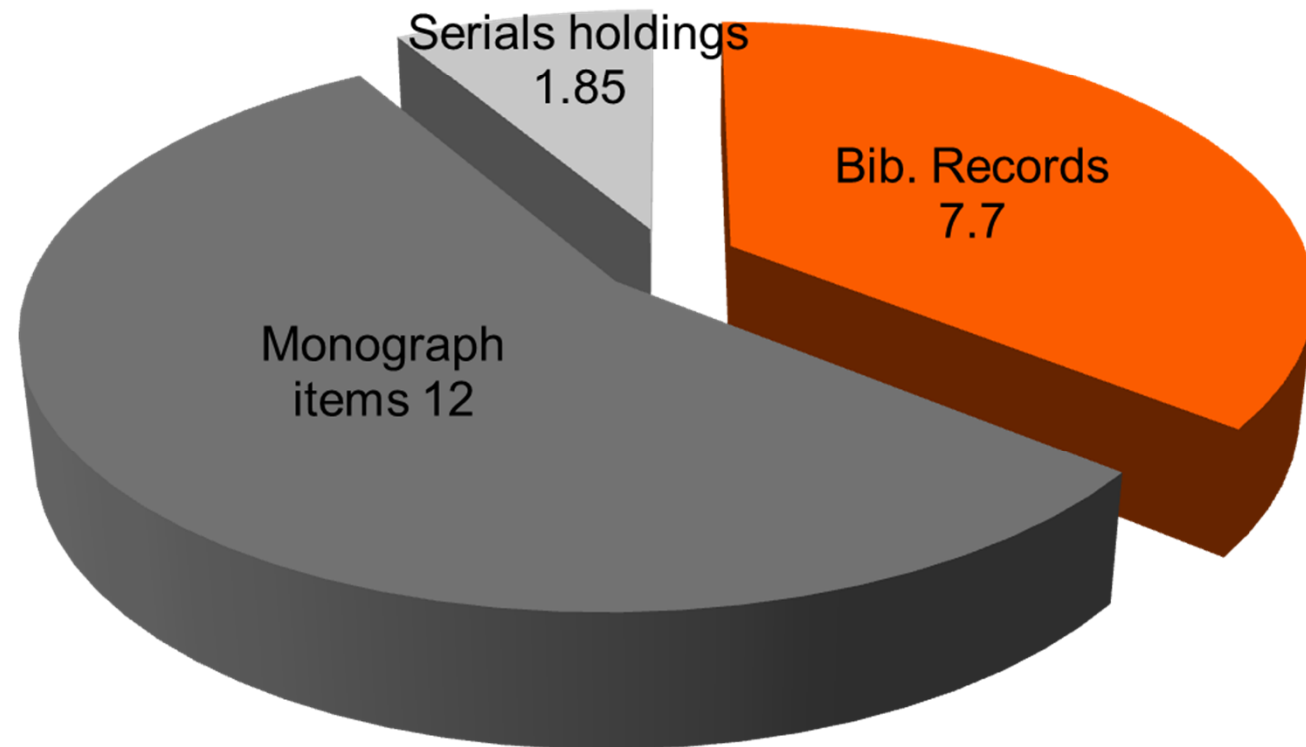
building a good **base Primo** implementation



What does it consist of?

Central Data – Aleph

Aleph Consortium database - data of 80 institutions (million records)



Central catalog contains the main data to build a Primo data source for any institution

- Aleph Publishing Mechanism (APM) well integrates extended data from authority and holdings records

→ just set it right up



A local ILS contains all Aleph data for the local Primo View

- Does all this data exist centrally?

NO!



How can it be solved?

- We bring missing local fields (classifications etc.) into the Central catalog in HOL-records (by means of automatic data replication)
- ILS records not presented centrally, institutions can publish locally



Two words about eDOC repository (since 2004)

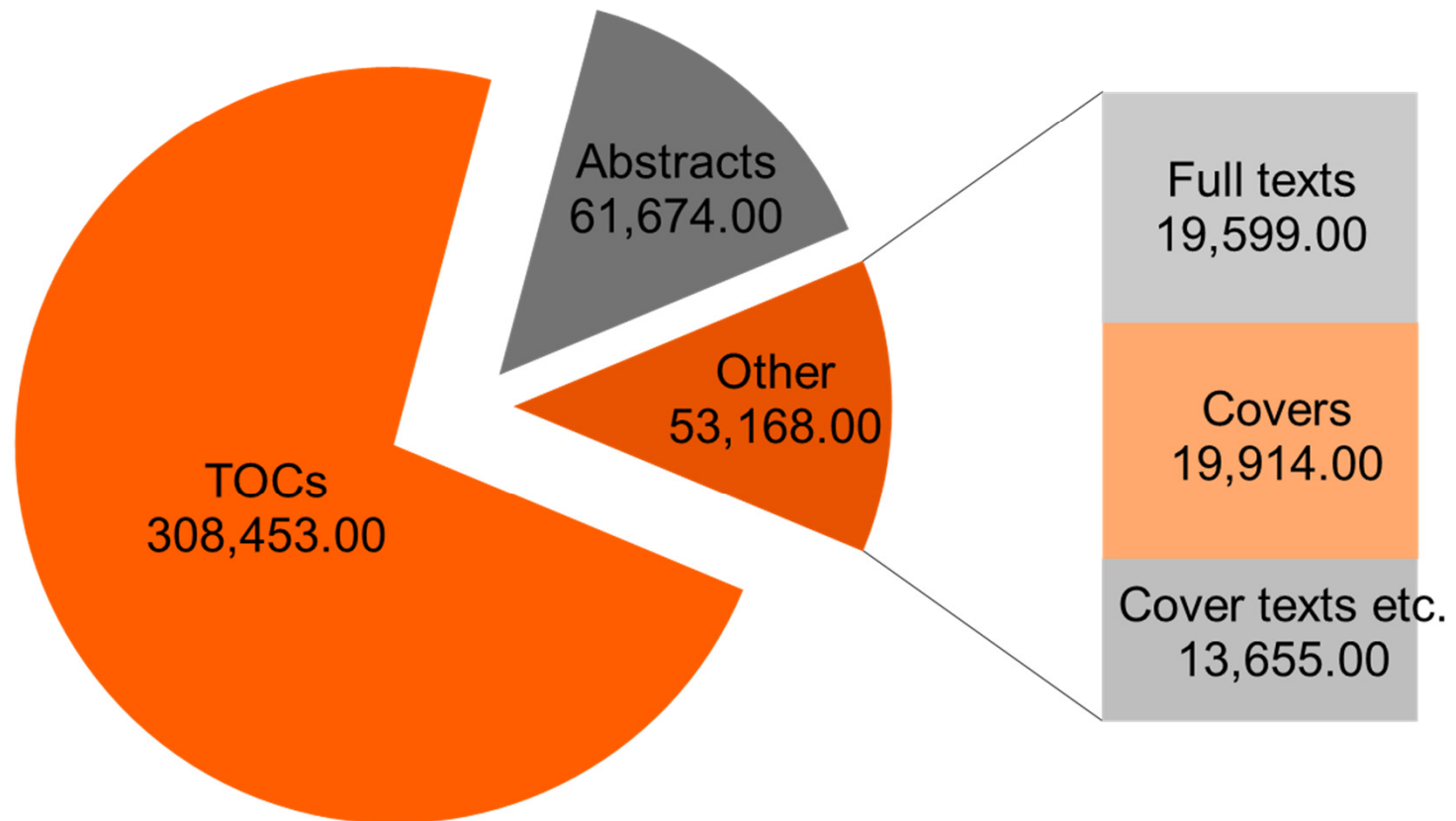
- In-house developed (MySQL – based)
- Primary tool for catalog enrichment projects
- eDOC updates Aleph with links & Aleph links its objects
- Objects can be stored in eDOC, at institutions or somewhere else

text contents of objects (out of *.pdf etc.) are all
stored in eDOC

→ a search engines-friendly repository ☺

Central Data – eDOC Repository

Central repository eDOC (total ~424.000 objects linked to Central catalog)



It is too little to setup Primo for Catalog data only

- We need fully integrate eDOC data and workflows into Primo
- Primo search engine can search text data and it should work for us from the start

We have data and experience of doing it with another search engine...



eDOC objects to be indexed in Primo

- TOCs
- Abstracts
- Full texts

they should also be searchable in separate Primo indexes!



And last but not least - full text updates

- Changes in eDOC should cause changes in Primo:
 - Updates
 - Deletions

Remember,
we are talking here about texts extracted from objects only!



Implement Standard Primo Functionalities

Implement Primo standard functions

We are consortium, and not all Primo features out-of-the-box work so good as for stand-alone installations!

*Note, we are talking here about Primo 2.1.x
(our production version is 2.1.10)*



Implement Standard Primo Functionalities

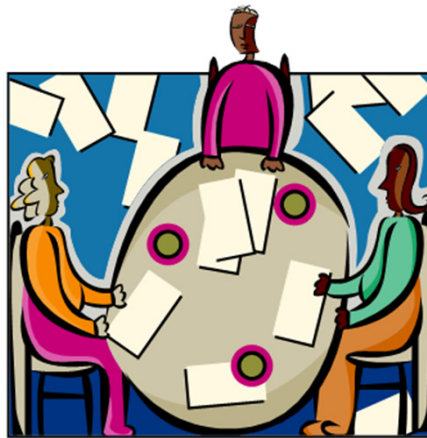
The essentials are:

- **Building Primo data sources**
 - out of Central data of consortium
 - Real Time Availability (RTA)
 - Full text indexing
- Work with Back Office for multi-institutions
- Front End parameterization
- Authentication (Shibboleth)

**”Building Primo data sources” –
became the **most critical** and **challenging**
item in the project, therefore it deserved a
special attention here!**

Implementing Primo – Issues and Problems

- Our Aleph consortium software is “made by Exl” but it does not fully suit Primo models



- Compromises and future solutions became our business

Implementation Issues – Central Data

- No efficient solution to produce Primo data sources for institutions from the Central Catalog

“existing options” bring **big data redundancies**



How would we want to have it?

Just to extract records of institution A from the central database!

What was possible out-of-the-box?

- Always load **all central records** into Primo database
- Then filter data out (e.g. via a view search scopes)

for us it would mean **23m** records instead of **12m** (in 2009)

- No efficient solution for institutions from “existing options”

How would it work?

Just a

How was it implemented?

- Always load data
- Then filter data

for us it would mean **2x**

- Primo data sources for long

redundancies



use

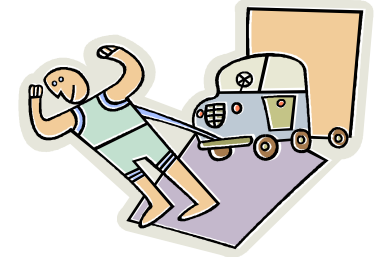
of **12m** (in 2009)

Implementation Issues – Central Data

- **No local bibl.-system numbers stored centrally**
 - mandatory for (**RTA**) implementation – a **critical local Primo function**
 - used to build back links to local OPACs
- **Too much administrative data stored centrally**
>1.500 lines in tab_sub_library.lng
 - it can only be partially interpreted by Primo and with high parameterization efforts



Challenges



Full text indexing: there are two options Primo offers:

- integrate texts into pnx records (A)
- write a customer Java plug-in specified by Exl (B)

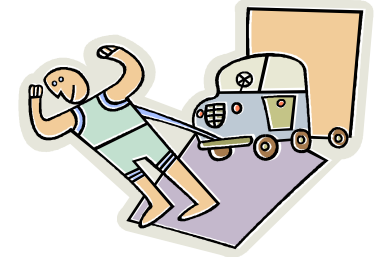
(A) – not practical, especially for repositories

(B) – may satisfy our requirements but ...

→ we would **be the 1st** who will do it ☹

→ Exl agreed on support if necessary ☺

Challenges



- **Some good catalog data loaded centrally and our libraries eager to have them locally**
 - 650.000 article records from 38.000 congress and serial titles
 - institutions have only main title records and wanted to search them by articles!

For OBVSG it would be a complex Aleph project

→ We should make Primo work for us and push these records into institutions views

Let's sum our problems up

Primo would do the job for us should we provide all that missing data!



How can we deal with this?

It seems that “help yourself” is our only choice...

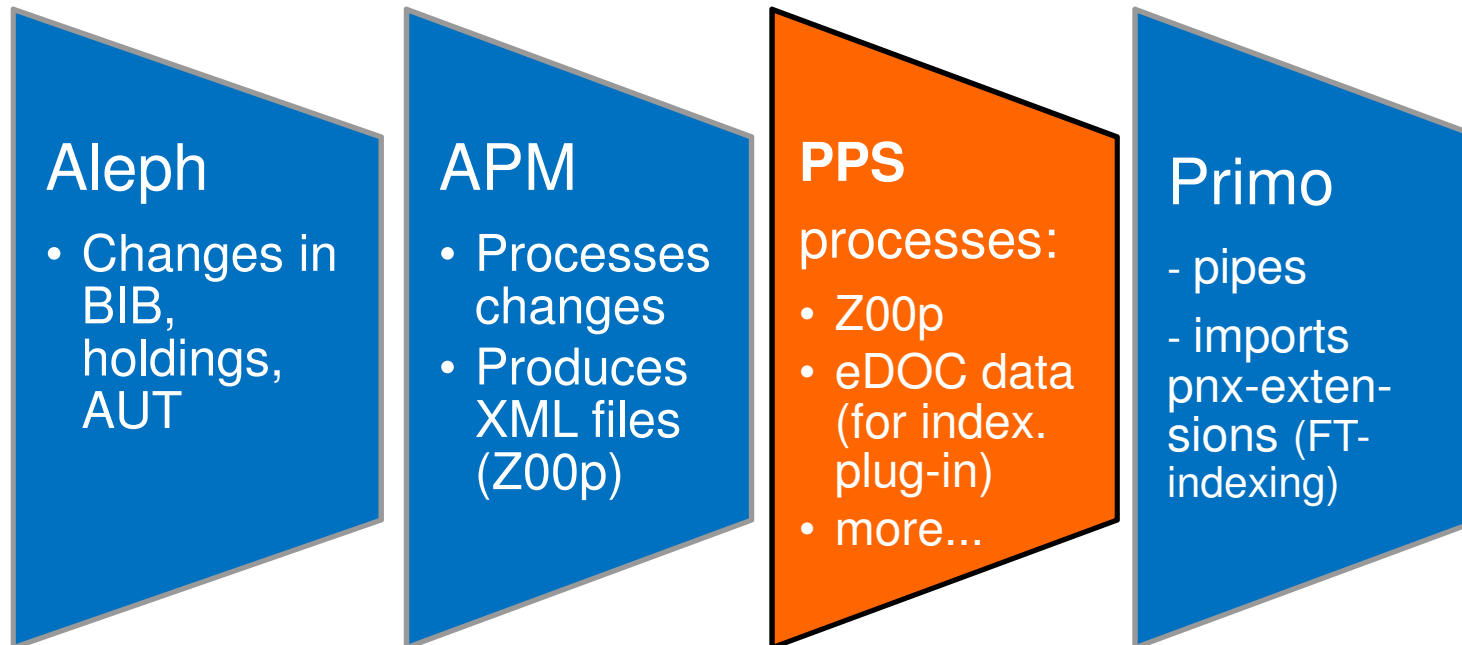
We need to implement a local data processing that:

- enriches data prepared by APM
- integrates eDOC
- implements indexing plug-in enabling Primo full text indexing
- runs efficiently and automatically (as it would be a part of standard Primo)

PPS – Primo Data Preparation System

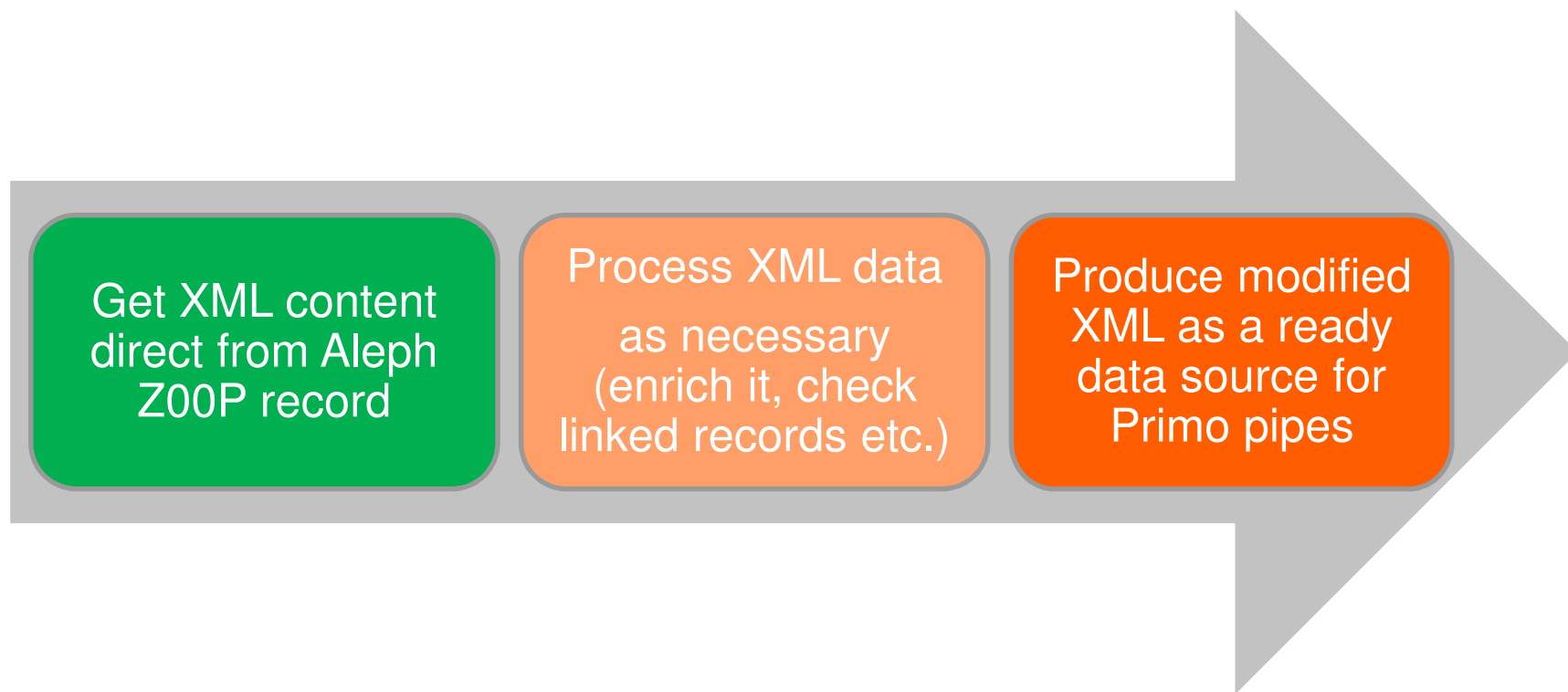
To solve our data problems we developed PPS

PPS is an application layer that runs after APM and before central Aleph Primo pipes



APM – Aleph Publishing Mechanism

PPS processing principles



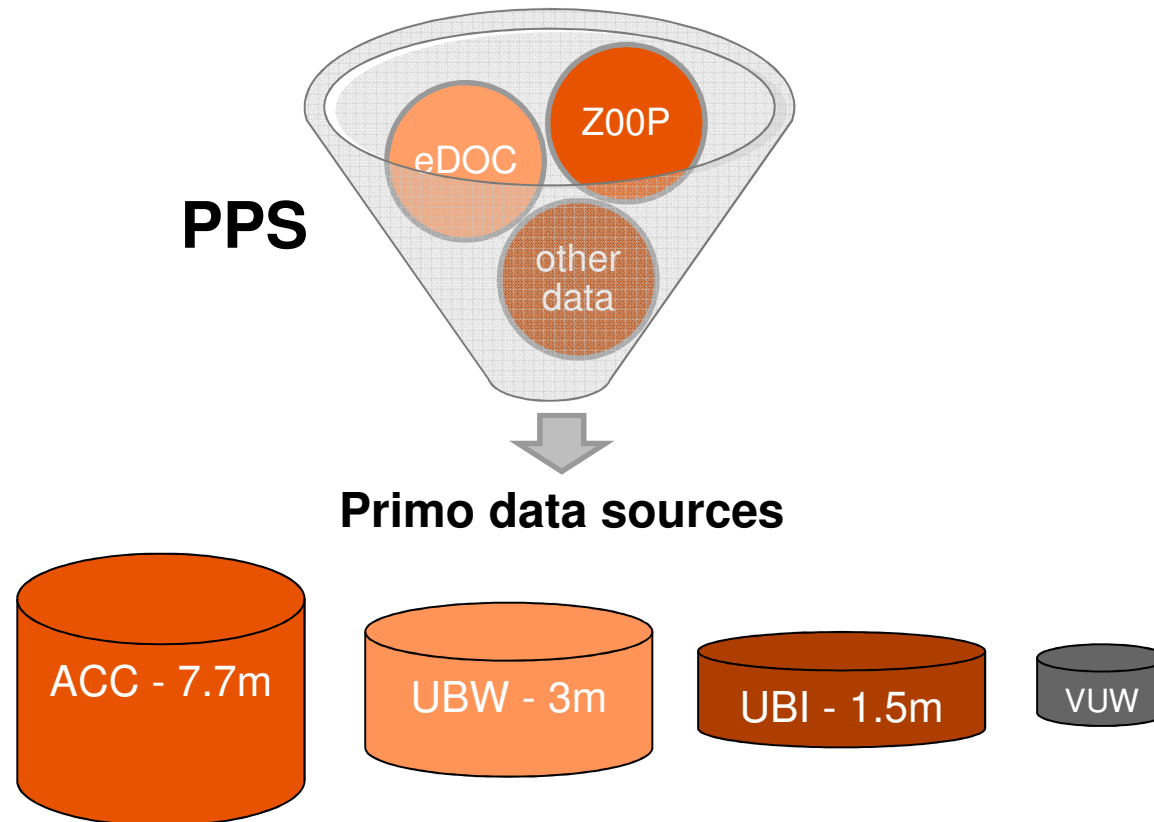
We began with PPS in 04.2009, productive it runs from 10.2009

.... PPS helped us indeed to solve the data problems!



PPS – Producing Primo Data Sources

From Z00P and eDOC PPS produces redundant-free Primo data sources for institutions



PPS – Producing Data Sources

Each institution receives the following data from the central sources:

- a copy of consortium record
- local fields (classification etc.) from central HOL records
- additional data (e.g. article records of serials/collections)
- eDOC objects
- local system numbers of bibl. records

All this, along with data from other local pipes (incl. SFX, ML etc.), provides a **solid basis** to build a **good local Primo view** in our consortium installation ☺



PPS – Demo: Title in Local Display Innsbruck

Search for **GO** >

Look for my query: ▼

[New Search](#) [Advanced Search](#)

**Sample
Data added by PPS
in local view**

Details

Result 1

**Cella : strutture di emarginazione e disciplinamento ;
Strukturen der Ausgrenzung und Disziplinierung ;
[Forschungs- und Ausstellungsprojekt des Instituts für
Kunstgeschichte der Universität Innsbruck im Rahmen des
Forschungsschwerpunktes "Politische Kommunikation und die
Macht der Bilder" ; Teil des Programms "Geschichte trifft
Zukunft 1809 - 2009" des Landes Tirol ; Ausstellung:
Complesso Monumentale di San Michele a Ripa, Rom ... 6.
November - 28. November 2009 ; Katalog]**

Author: [Bertsch, Christoph \[Hrsg.\]](#) ;
[Alge, Ingmar \[Ill.\]](#) ;
[Antenhofer, Christina](#)

Publisher: Innsbruck ; Wien [u.a.] ; Skarabaeus-Verl.

Creation Date: 2010

Format: 479 S. : zahlr. Ill., Kt.

Identifier: ISBN 978-3-7082-3268-3

Description: Literaturangaben

Description: Literaturangaben

Related Titles: Ausstellungskatalog / Institut für Kunstgeschichte der
Universität Innsbruck ; 26

Language: German ; Italian

Local information: Tir 480

local classifi-
cation (from HOL)

BK (Basis Classification): [20.00 - Kunstwissenschaften: Allgemeines](#)

Subjects: [Kunst / Gefängnis / Geschichte 1970-2009 / Ausstellung /
Innsbruck <2009>](#) ;
[Kunst / Ausgrenzung / Geschichte 1970-2009 / Ausstellung / Innsbruck
<2009>](#)

Downlink: [AC07950455](#)

eDOC object

[→ Inhaltsverzeichnis](#)

[→ Titel im Online-Katalog der Universität Innsbruck](#)

☐ Add to e-Shelf

Available at library ([GetIt](#))

link to articles

Back link to local OPAC

Real Time Availability

local system number

PPS – Demo: Title in Central Display

Details

[Back to results list](#)

Result 1

Cella : **strutture di emarginazione e disciplinamento** ; Strukturen der Ausgrenzung und Disziplinierung ; [Forschungs- und Ausstellungsprojekt des Instituts für Kunstgeschichte der Universität Innsbruck im Rahmen des Forschungsschwerpunktes "Politische Kommunikation und die Macht der Bilder" ; Teil des Programms "Geschichte trifft Zukunft 1809 - 2009" des Landes Tirol ; Ausstellung: Complesso Monumentale di San Michele a Ripa, Rom ... 6. November - 28. November 2009 ; Katalog]

Author: [Bertsch, Christoph \[Hrsg.\]](#) ; [Alge, Ingmar \[Ill.\]](#) ; [Antenhofer, Christina](#)

Subjects: [Kunst](#) ; [Gefängnis](#) ; [Geschichte 1970-2009](#) ; [Ausstellung](#) ; [Innsbruck <2009>](#) ; [Kunst](#) ; [Ausgrenzung](#) ; [Geschichte 1970-2009](#) ; [Ausstellung](#) ; [Innsbruck <2009>](#)

BK (Basis Classification): [20.00 - Kunstwissenschaften: Allgemeines](#)

Format: 479 S. ; zahlr. Ill., Kt.

Description: Literaturangaben

Related Titles: [Ausstellungskatalog / Institut für Kunstgeschichte der Universität Innsbruck](#) ; 26

Publisher: [Innsbruck](#) ; [Wien \[u.a.\]](#) ; [Skarabaeus-Verl.](#)

Creation Date: 2010

Language: German ; Italian

Identifier: [ISBN 978-3-7082-3268-3](#)

Unique consortium ID: [AC07950455](#)

Downlink: [AC07950455](#)



[Titel im Online-Katalog des OBV](#)



[Inhaltsverzeichnis](#)

☐ [Add to e-Shelf](#)

Availability and location:

- [Österr. Nationalbibliothek](#)
- [University of Innsbruck](#)

link to articles ("native")

link to eDOC (FT-search)

link to local OPACs via
added by PPS field **IDL**

**Sample
Data added by PPS
in Consortium view**

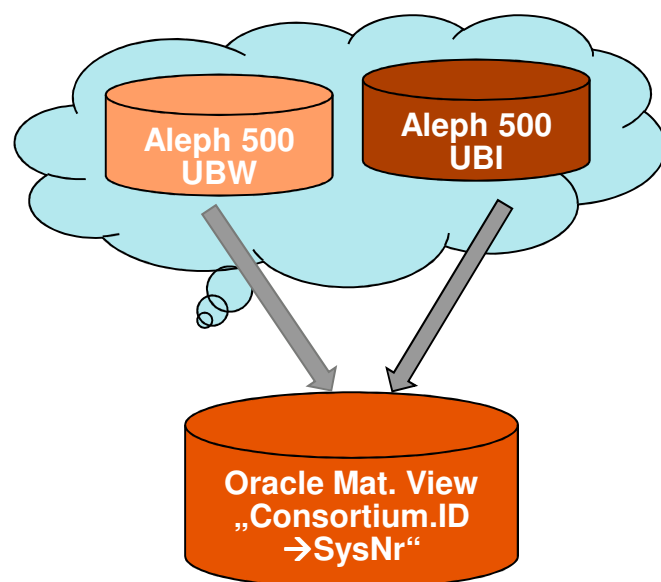
IDL

is the inst. code.
calc. by PPS,
added to MARC
XML from
items/holdings
fields after
processing of
tab_sub_library.

PPS – Behind the Scene: System Numbers

How local system numbers are obtained?

- PPS accesses direct indexes of local Aleph systems via SQL*NET
- for a higher efficiency, the extracted pairs Bib.ID→SysNr are stored centrally (as Oracle materialized views)



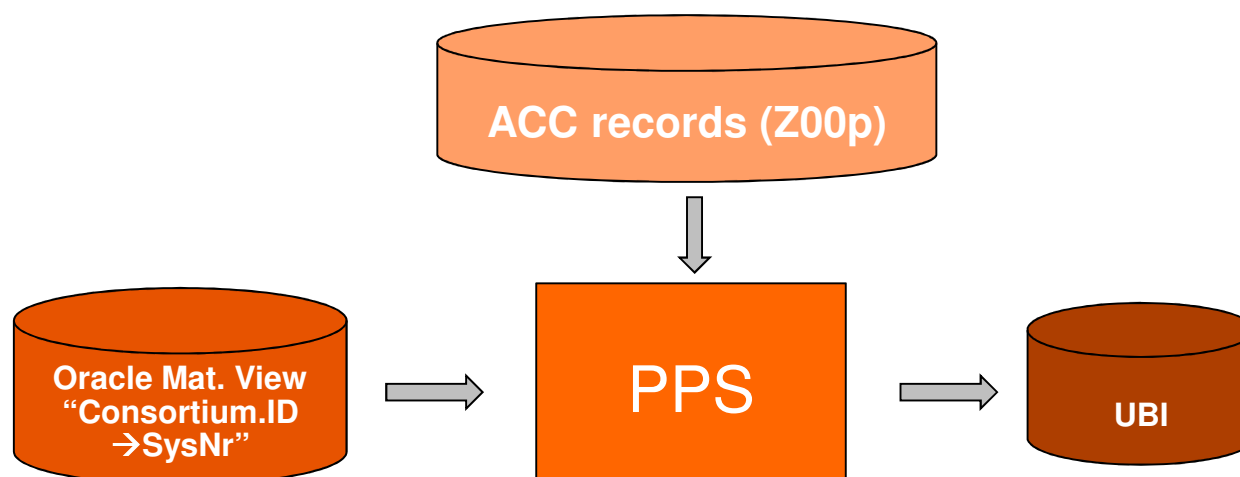
While building a data source for institution UBW PPS does:

- from each central record it takes Bibl.ID and checks it against the UBW part of table
- if ID found, then this central record is taken along with its local SysNr and written to the data source of UBW

PPS – Behind the Scene: Linking Articles

How central article records pushed to local views? Sample UBI.

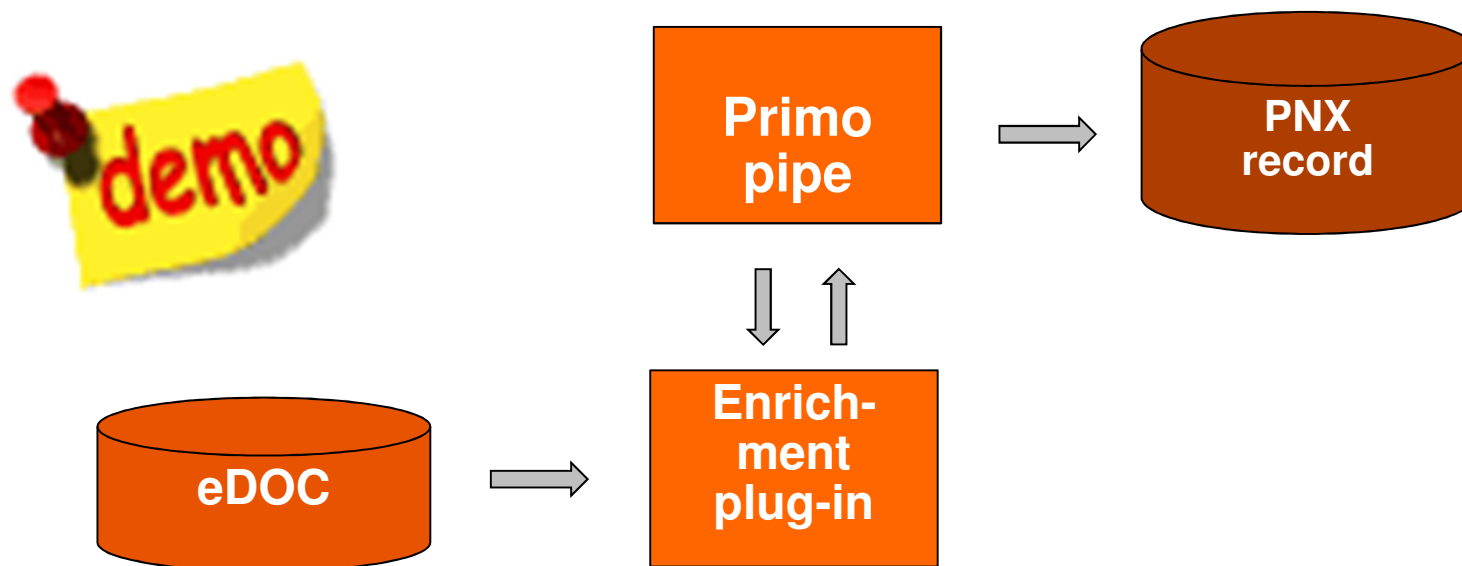
- If a processed record is article, its XML tag 599 contains Bibl.ID of the main record
- PPS uses this ID and accesses the index table of UBI
- If record found then PPS writes this article record into the data source of UBI



Changing Primo Full display – Enrichment Plug-In

We developed an enrichment plug-in manipulating data in PNX record

- Currently it “expands” texts of abstracts extracted from eDOC into the Primo full display
- The program accesses eDOC DB, fetches object and formats it as necessary for the display section of PNX record



Full Text Indexing Odyssey

It works now as we wanted, but the road was long and winding!



.... and we were pioneers of this road

Full Text Indexing Odyssey

We decided to develop indexing plug-in and transport text objects into Primo by means of Primo import tool

- the “import of pnx extensions” tool runs outside of Primo pipes
- each imported object is stored as an addition to pnx record – “pnx extension” – these both are then indexed
- the approach provides important decoupling of processing of bibliographic data and linked digital objects (both may have various update frequencies etc.)

Full Text Indexing Odyssey

On completion of plug-in in 10/2009, quite a lot of time and work was necessary to complete the planned implementation

- a good co-operation with Exl developers brought functional improvements to the original approach
- many problems and bugs were solved and fixed
- a stable solution came from Exl in April 2010
- quite a lot from our work Exl brought to Primo 3 ☺
- current solution may produce “redundant” data!



Additional Web Services

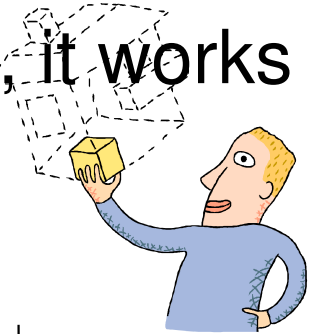
We also had a chance to integrate and develop the following web services in Primo

- Web 2.0 Features
 - tagging and reviews, social web and bookmarking
- Books Preview (Google books)
- Linking to Wikipedia (German)
- Location maps for libraries (Open street

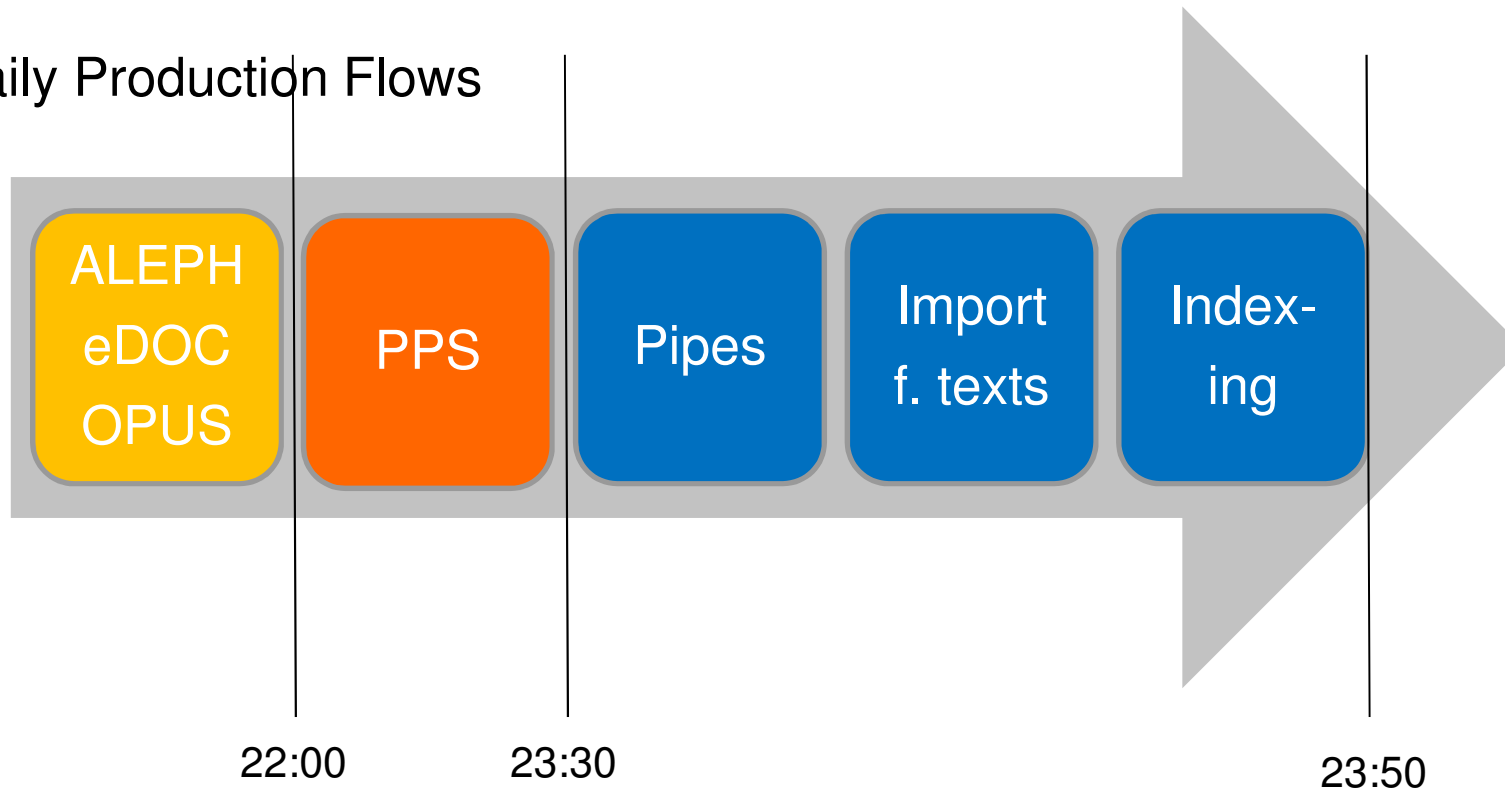


Production Workflows Overview 2010

We are happy and proud of our Primo Base, it works now **fully** as it should from 04.2010 ;-)

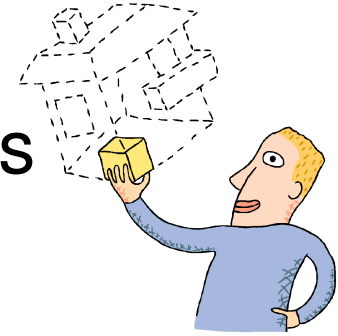


Daily Production Flows



Objectives 2009 - Base Primo Implementation

Having implemented PPS
and integrating it into Primo workflows
we have also met the following
three objectives:



- **Implement solutions and workflows enabling easy integration of new institutions**
- **High level of automation for all operations**
- **All three participants run Primo productive / Q.1 2010**



EXPERIENCES and TO DO

Experiences

Multi tenancy is even more important than expected; at the moment

- No free or at least restricted access to production server within the shared environment possible
- Transfer of data from test to production server is difficult
- Problems with data export of a view
- Parameterization of Deduplication and FRBR act on the whole instance

Primo Version 3 and even more the new 2011 Primo data model is highly awaited from Ex Libris

To Do

High availability is crucial for this architecture

- Still unsolved problems with Oracle RAC (Real Time Application Cluster)
- “Certified” by Ex Libris does not mean, that it is always working
 - We found failover problems with Primo itself
- Testing RAC/Primo in cooperation with Ex Libris
- At the moment Ex Libris uses not thoroughly tested versions of Oracle (RAC)
 - In professional training units there was strong advice, to use sufficiently patched versions only
 - Four out of five tests tried by our system administrators failed ...

To Do

Better use of virtualisation

- Unnecessary hardware has high impact on project costs
- Only 5% of search machine infrastructure used at the moment
- We had to “force” Ex Libris that we can use virtualisation via KVM on our test machines → it works smoothly
- Migration to Primo Version 3 is pending – is there really a reason to buy new machines for this?
- As far as we know, the hype technology “cloud” depends fundamentally on virtualisation ...

We will address and solve this issues in cooperation with Ex Libris!

EFFORTS and COSTS

Efforts: Implementation

Principal architecture and technical infrastructure

- Duration: about 1 year
- Core work has been done by a five members team of OBVSG
 - 2 system librarians
 - 2 analysts / programmers
 - 1 system administrator
- More servers needed than we had expected

Efforts: Implementation

Front End interfaces and Shibboleth

- Duration: several months
- up to 3 full time system librarians at the University of Vienna and at the University of Innsbruck

Data manipulations

- Checks for and corrections of incorrect or incomplete data in the central catalogue
- Standardisation of holdings data elements

Efforts: Implementation

The project implementation would not be possible without excellent co-operation with participating institutions and support and solutions provided by Ex Libris.

On-going costs

Complex system of high quality needs an appropriate amount of resources

- Developing of standard procedures and homogeneous data situation makes the situation manageable
- Start efforts can be split about the implementers
- Newly entering institutions can strongly benefit from the work done

On-going costs

Cost situation

- FTE 1: fulltime equivalent of qualified IT staff
- FTE 2: fulltime equivalent of high qualified library staff
- No licences considered
- Hardware, IT infrastructure, system operation and data preparation included

	FTE 1	FTE 2
Small institution	23%	29%
Medium institution	61%	77%
Big institution	98%	123%
Very big institution	117%	146%

NEXT STEPS

Next steps

- Migration to Primo Version 3
- Evaluation of Primo Central
- New participants
- Further improve quality and completeness of data and service levels

CONCLUSION

Conclusion

Implementation

- Our biggest project since introduction of Aleph
- Went to full production reaching main goals
- Established intense cooperation between Ex Libris, OBVSG and the Universities of Vienna and Innsbruck

Conclusion

Consortium benefits

- Concentration of resources and know how to form a critical mass
- Preparation of standard solutions, which can be further tailored to fulfil individual needs
- High service level at competitive costs

Conclusion

Future

- We are looking forward to the announced improvements
- We are confident to increase the service quality even more
- Looking forward to the substitution of the Aleph OPAC and coverage of journal articles
- Discovering still new and fascinating possibilities ...

Thanks!

THANK YOU FOR YOUR PATIENCE!

Wolfgang.Hamedinger@obvsg.at

Victor.Babitchev@obvsg.at

The Austrian Library Network

Visit our Primo views:

<http://usearch.univie.ac.at/>

- Univ. Vienna

<http://search.obvsg.at/UIB>

- Univ. Innsbruck

<http://search.obvsg.at/UVW>

- Vet.-med. Univ. Vienna

<http://search.obvsg.at/ACC>

- the Consortium view