

A Dataspace Odyssey: The iMeMex Personal DataSpace Management System

Lukas Blunschi, Jens Dittrich, Olivier Girard,
Shant Karakashian, Marcos Salles

ETH Zurich, iMeMex.org



What is a DataSpace Management System?

- Vision paper by Mike Franklin, Alon Halevy and David Maier
From Databases to Dataspaces: a New Abstraction for Information Management.
SIGMOD Record, December 2005
- Idea: ALL data belonging to a particular person, task, or company gets assigned to a logical dataspace, no upfront investments for schemas.
- Dataspace independent from formats, devices, locations, and systems owning the data.
- What we need is a new kind of system able to manage dataspace across devices and formats.

From Database Systems to DataSpace Systems

- DataSpace management systems can be seen as a hybrid of
 - a search engine,
 - a databases management system,
 - an information integration system,
 - and a data sharing system.
- iMeMex is one of the first DataSpace Management Systems
- iMeMex is designed as a highly extensible framework (Eclipse-like, OSGi service-oriented architecture)
- Current focus of iMeMex: personal dataspace
- However, we are also targeting other applications including scientific dataspace.

iMeMex Project: Integrated MeMex

- Project started 2 years ago
- Our goal: Build the first open source dataspace management system
- All Java, service-oriented (OSGi, cmp. Eclipse project)
- Plus platform specific extensions in C (currently for Windows and Mac)
- iMeMex is an extensible framework for dataspace management
- Supported by SNF (Swiss National Science Foundation)
- Demo at VLDB '05, papers at VLDB '06 and SIGIR PIM '06
- State of the system, milestones:
 - December '06: 0.42.0, RC1a developer version released/open source ✓
 - Beg of Feb '07: 0.43.0, RC1b developer version
 - End of Feb '07: 0.43.x, RC2 developer version + RC1 AJAX client
 - End of May '07: 0.44.x, RC3 developer version + RC2 AJAX client
 - End of July '07: 0.45.x, RC1 end-user version

Literature on PIM and Dataspaces

- J. Dittrich, M. Salles, D. Kossmann, L. Blunschi
iMeMex: Escapes from the Personal Information Jungle (Demo Paper)
In VLDB 2005, September 2005.
- M. Franklin, A. Halevy, D. Maier
From Databases to Dataspaces: A New Abstraction for Information Management
SIGMOD Record, 34(4):27–33, December 2005.
- J. Dittrich
iMeMex: A Platform for Personal DataSpace Management
SIGIR PIM 2006, August 2006.
- J. Dittrich, M. Salles
iDM: A Unified and Versatile Data Model for Personal Dataspace Management
VLDB 2006 (IIS Track): September 2006.
- L. Blunschi, J. Dittrich, O. Girad, S. Karakashian, M. Salles
A Dataspace Odyssey: The iMeMex Personal Dataspace Management System
CIDR 2007 (Biennial Conference on Innovative Data Systems Research), Jan. 2007
- J. Dittrich, L. Blunschi, O. Girad, S. Karakashian, M. Salles, M. Färber
From Personal Desktops to Personal Dataspaces: A Report on Building the iMeMex
Personal Dataspace Management System
BTW 2007 (Biennial German Database Conference), March 2007

Demo Screenshots

iMeMex AJAX Interface

search&query box

Administration About Help

Search&Query: Find

registered data sources

- ▶ **inbox.misc2 email (1)**
- ▶ testData (2)
- ▶ atom feed (3)
- ▶ rss feed (4)
- ▶ IMEMEXTEST database (5)

inbox.misc2 email More Info

fullname	inbox.misc2
unreadMessages	0
exists	1
OID	1
deletedMessages	0
class	emailfolder
messages	12
newMessages	0
uri	imap://shrek04.inf.ethz.ch
subscribed	1

imap://shrek04.inf.ethz.ch
Explore Context: [Connections](#)

5 resource views found.

Transferring data from localhost...

NOTE: we did not write any HTML/Javascript to generate this AJAX application. Our AJAX application was automatically generated from a Java program using GWT.

IMAP Mail Data Source

The screenshot shows the iMeMex Dataspace Navigator interface. The browser address bar displays `http://localhost:8050/GWTClient.html`. The navigation menu includes "Administration", "About", and "Help". A search bar is labeled "Search&Query:" with a "Find" button. The main content area is divided into two panes. The left pane shows a tree view of data sources, with "inbox.misc2 email (1)" expanded. A red box highlights this section, and a red arrow points from the text "email messages" to it. The right pane displays the details of a selected email message:

Fwd: IASTED PCDN 2007 Innsbruck [More Info](#)

to	all
from	Peter Arbenz
received	03/01/07 09:01:54
OID	41
SEEN	1
sent	03/08/06 11:19:11
subject	Fwd: IASTED PCDN 2007 Innsbruck
class	emailmessage
size	271527
contentType	multipart/MIXED; boundary="-----=_NextPart_001_01C6C1EF.217FB580"
lastmodified	03/08/06 11:19:11
mimeType	multipart/mixed
messageID	<200608171319.11537.arbenz@inf.ethz.ch>
uri	imap://shrek04.inf.ethz.ch/10

Below the message details, the URI `imap://shrek04.inf.ethz.ch/10` is shown in green. A button labeled "Explore Context: Connections" is visible. At the bottom of the interface, it states "5 resource views found." and "Transferring data from localhost..."

email messages

File System Data Source

The screenshot shows the iMeMex Dataspace Navigator interface. The browser address bar displays `http://localhost:8050/GWTClient.html`. The navigation pane on the left shows a tree structure of resources. A red box highlights the `testData (2)` folder, which contains several sub-items, including `test.xml`. A red arrow points from the `test.xml` entry in the tree to the right-hand pane, which displays detailed metadata for this file.

test.xml More Info

readonly	0
OID	6985
size	512
class	xmlelem
hidden	0
lastmodified	04/02/06 19:59:26
uri	file:///test.xml

<file:///test.xml>

Explore Context: [Connections](#) [Same Time](#) [Same Size](#)

6 resource views found.

Transferring data from localhost...

file system data

ATOM Feed Data Source

The screenshot shows the iMeMex Dataspace Navigator interface. The browser address bar displays `http://localhost:8050/GWTClient.html`. The main content area is divided into a left sidebar and a right pane. The sidebar shows a tree view of data sources: `inbox.misc2 email (1)`, `testData (2)`, and `atom feed (3)`. The `atom feed (3)` folder is expanded, showing a list of news messages. The first message, `Schwarzenegger's Leg Surgery Successful - MyFox Washington DC`, is highlighted with a red box. The right pane displays the details for this selected message, including its `OID`, `id`, `created`, `modified`, `issued`, `title`, `atom entry class`, and `uri`. A red arrow points from the text `ATOM news messages` to the list of messages in the sidebar.

Administration About Help

Search&Query: Find

iMeMex.org
Personal DataSpace Management System

- ▶ inbox.misc2 email (1)
- ▶ testData (2)
- ▼ atom feed (3)
 - ▶ **Schwarzenegger's Leg Surgery Successful - MyFox Washington DC**
 - ▶ Court: Execute Saddam Within 3
 - ▶ Islamic forces retreat in Somalia
 - ▶ Gas Pipeline Blast Kills 260 in Nic
 - ▶ Veil Slowly Lifting Over Castro M
 - ▶ Biden urges Bush to 'level' with p
 - ▶ Thieves Rob Church During Chris
 - ▶ Reps spar over Quran usage - N
 - ▶ Visa USA sees disappointing finis
 - ▶ Oil Falls As Warm Weather Over
 - ▶ Azerbaijan will not buy Russian g
 - ▶ Security flaw found in Microsoft's
 - ▶ Analysts are not optimistic about
 - ▶ Wikipedia Founder Plans Search
 - ▶ Garcia leads Eagles over Cowboy
 - ▶ Jets, Dolphins Scoreless at Halfti
 - ▶ Lakers' Christmas spirit gets sap
 - ▶ Fans Pay Tribute to the Late Jam

Schwarzenegger's Leg Surgery Successful - MyFox Washington DC More Info

OID	16513
id	tag:news.google.com,2005:cluster=424b3a6b
created	
modified	2006-12-26T20:20:00+00:00
issued	2006-12-26T20:20:00+00:00
title	Schwarzenegger's Leg Surgery Successful - MyFox Washington DC
atom entry class	class
uri	http://shrek04.inf.ethz.ch/3 ./my:feed/my:entry[1]

[http://shrek04.inf.ethz.ch/3|./my:feed/my:entry\[1\]](http://shrek04.inf.ethz.ch/3|./my:feed/my:entry[1])

5 resource views found.

Transferring data from localhost...

Oracle Database Data Source

Administration About Help

Search&Query: Find

inbox.misc2 email (1)
testData (2)
atom feed (3)
rss feed (4)
IMEMEXTEST database (5)
 ADDRESSES table
 pk: 2
 pk: 3
 pk: 55
 CUSTOMERS table
 ORDERS table

pk: 55 More Info

OID	16554
ID	55
ZIPCODE	8000
class	sql row
ID2	3
LASTMODIFIED	04/10/06 11:59:17
STREET	blu
uri	jdbc:///ADDRESSES/pk: 55
LOCATION	blu

jdbc:///ADDRESSES/pk: 55

20 resource views found.

Transferring data from localhost...

Database data source: either warehoused or mediated

Keyword Search

The screenshot shows the iMeMex Dataspace Navigator interface. The browser address bar displays `http://localhost:8050/GWTClient.html#HTMLPanel_394`. The search bar contains the query `halevy`. The left sidebar shows a navigation tree with the following items:

- inbox.misc2 email (1)
- testData (2)
- atom feed (3)
- rss feed (4)
- IMEMEXTEST database (5)

The main content area displays search results for the query 'halevy'. The results are as follows:

- M. Franklin and A. Halevy and D. Maier. From Databases to Dataspaces: A New Abstraction for Information Management. 2005.** [More Info](#)
`file:///CIDR2007/Marcos' documents/papers/imemex.bib|bibtex|FHM05`
- X. Dong and A. Halevy and J. Madhavan and E. Nemes. Reference Reconciliation in Complex Information Spaces. 2005.** [More Info](#)
`file:///CIDR2007/Marcos' documents/papers/imemex.bib|bibtex|DHMN05`
- Xin Dong and Alon Y. Halevy and Ema Nemes and Stephan B. Sigundsson and Pedro Domingos. SEMEX: Toward On-the-fly Personal Information Integration. 2004.** [More Info](#)
`file:///CIDR2007/Marcos' documents/papers/imemex.bib|bibtex|DHN+04`
- X. Dong and A. Halevy. A Platform for Personal Information Management and Integration. 2005.** [More Info](#)
`file:///CIDR2007/Marcos' documents/papers/imemex.bib|bibtex|DH05`
- A. Halevy and others. Crossing the Structure Chasm.** [More Info](#)

At the bottom of the results, it states: `20 resource views found.`

The status bar at the bottom left indicates: `Transferring data from localhost...`

Search&Query Granularity: Resource Views

Administration About Help

Search&Query: Find

iMeMex.org
Personal DataSpace Management System

- inbox.misc2 email (1)
- testData (2)
- atom feed (3)
- rss feed (4)
- IMEMEXTEST database (5)

M. Franklin and A. Halevy and D. Maier. From Databases to Dataspaces: A New Abstraction for Information Management. 2005. More Info
`file:///CIDR2007/Marcos' documents/papers/imemex.bib|bibtex|FHM05`

X. Dong and A. Halevy and J. Madhavan and E. Nemes. Reference Reconciliation in Complex Information Spaces. 2005. More Info
`file:///CIDR2007/Marcos' documents/papers/imemex.bib|bibtex|DHMN05`

Xin Dong and Alon Y. Halevy and Ema Nemes and Stephan B. Sigundsson and Pedro Domingos. SEMEX: Toward On-the-fly Personal Information Integration. 2004. More Info
`file:///CIDR2007/Marcos' documents/papers/imemex.bib|bibtex|DHN+04`

X. Dong and A. Halevy. A Platform for Personal Information Management and Integration. 2005. More Info
`file:///CIDR2007/Marcos' documents/papers/imemex.bib|bibtex|DH05`

A. Halevy and others. Crossing the Structure Chasm. More Info
20 resource views found.

Transferring data from localhost...

search&query granularity:
fine granular resource views
- not files or documents
Here:
entries from a BibTeX file

Search&Query Granularity: Resource Views

The screenshot shows the iMeMex Dataspace Navigator interface. The search query is 'halevy'. The left sidebar shows a tree view of resources: inbox.misc2 email (1), testData (2), atom feed (3), rss feed (4), and IMEMEXTEST database (5). The main content area displays search results. The first result is expanded, showing detailed metadata for the paper 'M. Franklin and A. Halevy and D. Maier. From Databases to Dataspaces: A New Abstraction for Information Management. 2005.' The metadata is as follows:

OID	577
id	FHM05
pages	27--33
year	2005
title	From Databases to Dataspaces: A New Abstraction for Information Management
class	bibtex_entry_article
*	FHM05bibtex_entry_articleM. Franklin and A. Halevy and D. MaierFrom Databases to Dataspaces: A New Abstraction for Information Management27--33SIGMOD Record577
author	M. Franklin and A. Halevy and D. Maier
volume	34
uri	file:///CIDR2007/Marcos' documents/papers/imemex.bib bibtex FHM05
journal	SIGMOD Record
number	4

The expanded view also shows the URI: `file:///CIDR2007/Marcos' documents/papers/imemex.bib|bibtex|FHM05`. Below this, the next result is visible: 'X. Dong and A. Halevy and J. Madhavan and E. Nemes. Reference Reconciliation in Complex Information Spaces. 2005.' with URI: `file:///CIDR2007/Marcos' documents/papers/imemex.bib|bibtex|DHMN05`. The bottom of the window shows 'Xin Dong and Alon Y. Halevy and Ema Nemes and 20 resource views found.'

expanded view of a result

Navigate Inside Files

The screenshot shows the iMeMex Dataspace Navigator interface. The left sidebar displays a file tree structure:

- textfiles
- TestFolder
- CIDR2007
 - Marcos' documents
 - papers
 - VLDB 2005 demo.tex
 - LaTeX**
 - \documentclass {vl}
 - abstract
 - Introduction
 - Related Work
 - iMeMex: As We May
 - Description of the D
 - Conclusions
 - CIDR 2007 demo.tex
 - VLDB 2006.tex
 - CIDR 2007.tex
 - imemex.bib
 - teaching

The right pane displays the details for the selected LaTeX file:

LaTeX [More Info](#)

OID 313
class texttext
uri file:///CIDR2007/Marcos' documents/papers/VLDB 2005 demo.tex|texDoc
file:///CIDR2007/Marcos' documents/papers/VLDB 2005 demo.tex|texDoc
Explore Context: [Connections](#)

A red box highlights the LaTeX file structure in the sidebar, and a red arrow points from this box to the text "structure found inside a LaTeX file".

6 resource views found.

Transferring data from localhost...

Structural Search (NEXI-like)

The screenshot shows the iMeMex Dataspace Navigator web application. The browser address bar displays `http://localhost:8050/GWTClient.html`. The search bar contains the query `//sigmod*//exper*`, which is highlighted with a red box. A red arrow points from this query to the explanatory text on the left. The left sidebar shows a tree view of the data space with folders like `inbox.misc2_email (1)`, `testData (2)`, `atom feed (3)`, `rss feed (4)`, and `IMEMEXTEST database (5)`. The main content area displays six search results, each with a title, a file path, and an 'Explore Context: Connections' button. The results are:
1. Experiment~1: Response Time on File I/O (file:///texFiles/sigmod2006/experiments.tex|texDoc0.4)
2. Experiment~2: Query Processing (file:///texFiles/sigmod2006/experiments.tex|texDoc0.5)
3. experiments.tex (file:///TexFileCollection/sigmod2006/experiments.tex)
4. Experiment~1: Response Time on File I/O (file:///TexFileCollection/sigmod2006/experiments.tex|texDoc0.4)
5. Experiment~2: Query Processing (file:///TexFileCollection/sigmod2006/experiments.tex|texDoc0.5)
6. experiments.tex (file:///TexFileCollection/sigmod2006/experiments.tex)
The status bar at the bottom indicates '6 resource views found.'

structural queries beyond tcsh,
similar to NEXI
(Narrowed Extended XPath I)

Search History

The screenshot shows the iMeMex Dataspace Navigator web application. The browser address bar displays `http://localhost:8050/GWTClient.html`. The search bar contains the query `//sigmod*//exper*`. A dropdown menu below the search bar shows a list of search history entries: `halevy`, `schwa`, `//sigmod*//exper*`, and `semes*/*[imemex]`. A red box highlights this dropdown menu, and a red arrow points from the text "search history" to it. The search results on the right side of the page include:

- Experiment~1: Response Time on File I/O** (More Info)
 - file:///sigmod2006/experiments.tex|texDoc0.4
 - Explore Context: [Connections](#)
- Experiment~2: Query Processing** (More Info)
 - file:///texFiles/sigmod2006/experiments.tex|texDoc0.5
 - Explore Context: [Connections](#)
- experiments.tex** (More Info)
 - file:///TexFileCollection/sigmod2006/experiments.tex
 - Explore Context:

Connections	Same Content Type	Same Time	Same Size
- Experiment~1: Response Time on File I/O** (More Info)
 - file:///TexFileCollection/sigmod2006/experiments.tex|texDoc0.4
 - Explore Context: [Connections](#)
- Experiment~2: Query Processing** (More Info)
 - 6 resource views found.

The status bar at the bottom indicates "Transferring data from localhost..."

search history

Keyword Prefix Search

The screenshot shows the iMeMex Dataspace Navigator interface. The search bar contains the text "schwa". The left sidebar shows a tree view of resources, with "rss feed (4)" highlighted. The main content area displays two search results, both titled "Schwarzenegger's Leg Surgery Successful - MyFox Washington DC". The first result has a URL starting with "http://shrek04.inf.ethz.ch/3|./my/feed/my:entry[1]", and the second result has a URL starting with "file:///xml/atom-example.xml|./my/feed/my:entry[1]". A red arrow points from the search bar to the first result. At the bottom of the window, it says "2 resource views found." and "Transferring data from localhost..."

prefix "schwa" will also return results with "Schwarzenegger"

Popups for Graph Navigation

The screenshot displays the iMeMex Dataspace Navigator web application. The browser address bar shows the URL `http://localhost:8050/GWTClient.html#con311`. The main interface includes a search bar with the query `halevy` and a file tree on the left. The file tree shows a directory structure with folders like `CIDR2007` and `Marcos' documents`, and files like `VLDB 2005 demo.tex`. A popup window titled `Connections` is open, showing a subgraph of the selected file. The popup contains a tree view with nodes like `VLDB 2005 demo.tex (2)`, `LaTeX`, and `abstract`. A red box highlights the popup window, and a red arrow points to it with the text `popups allow navigation on subgraphs`.

Administration About Help

Search&Query: halevy Find

iMeMex.org
Personal DataSpace Management System

abstract More Info

OID 322

class environment

uri file:///CIDR2007/Marcos' documents/papers/VLDB 2005 demo.tex|texDoc1

file:///CIDR2007/Marcos' documents/papers/VLDB 2005 demo.tex|texDoc1

Explore Context: Connections

Connections

- VLDB 2005 demo.tex (2)
 - LaTeX
 - \documentclass {vld...
 - abstract**
 - Modern computer work...
 - Introduction
 - Related Work
 - iMeMex: As We May Store
 - Description of the Demo
 - Conclusions

close

Transferring data from localhost...

Upcoming Features

- Declarative Pay-as-you-go Information Integration Framework
- Data Sharing Framework using distributed iMeMex instances
- Cost-based optimization
- <your plugin here>
- Details:
see our web-site: <http://www.imemex.org>