



Cross-Media Publishing Using XML Databases



- XQuery on the Web
- The XRX Ecosystem
- XRX Beyond the Web
 - Case Study 1: Generating an ebook on the fly
 - Case Study 2: A Complex publication project



XQuery on the Web



When I started implementing
XQuery back in 2003/4 ...



I was not convinced it would be
useful, sorry!



We already had XPath + XSLT.
What else did we need?



But then I found I could write
entire web applications in this
functional language!



Stick to what you know best:
XML



Develop your applications using nothing but XML and related standards

- (X)Query
- (R)EST
- (X)Forms

Filling the Gaps



Ok, so I understand I can use XQuery instead of PHP, but ...

Filling the Gaps

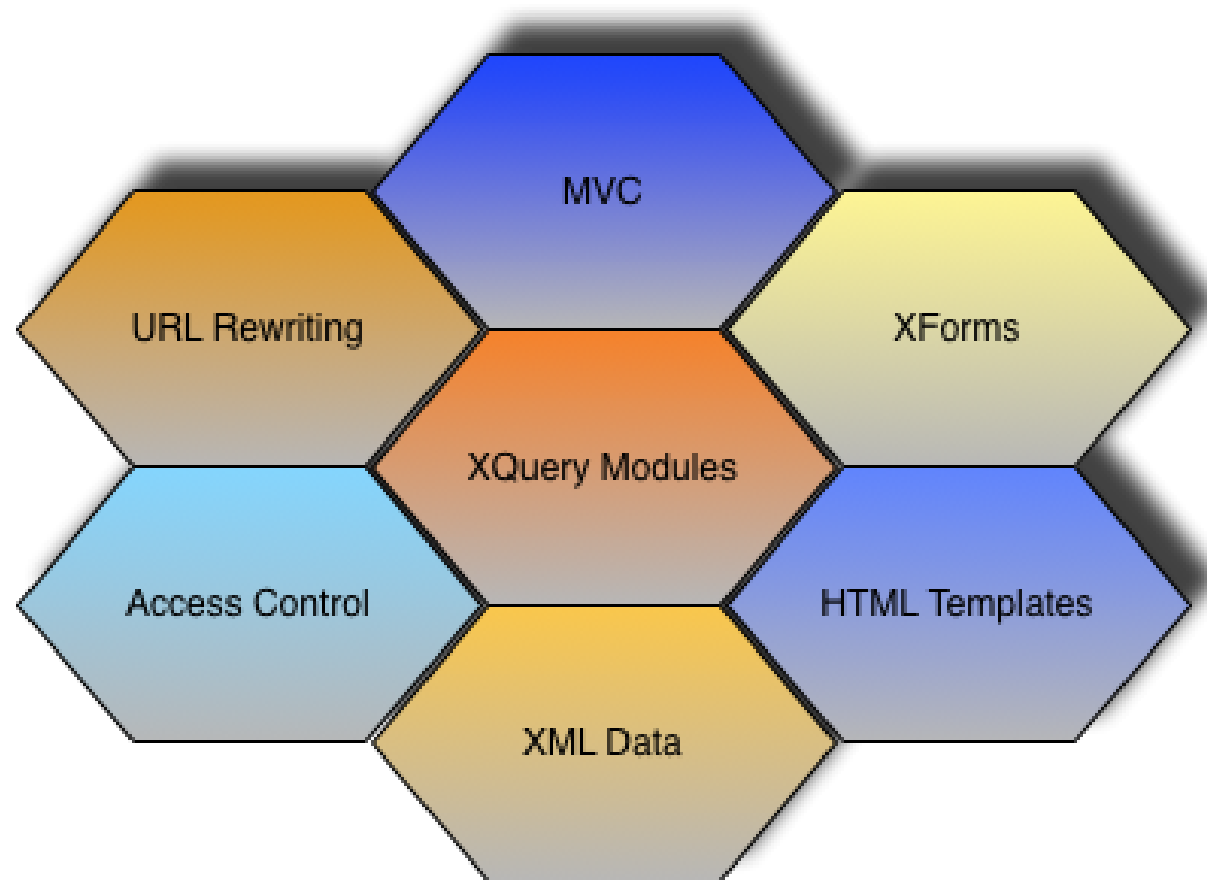


what about all the cool
frameworks and libraries?



The XRX Ecosystem

Application Components





XML DBs like eXist today provide an entire ecosystem, comprising things like

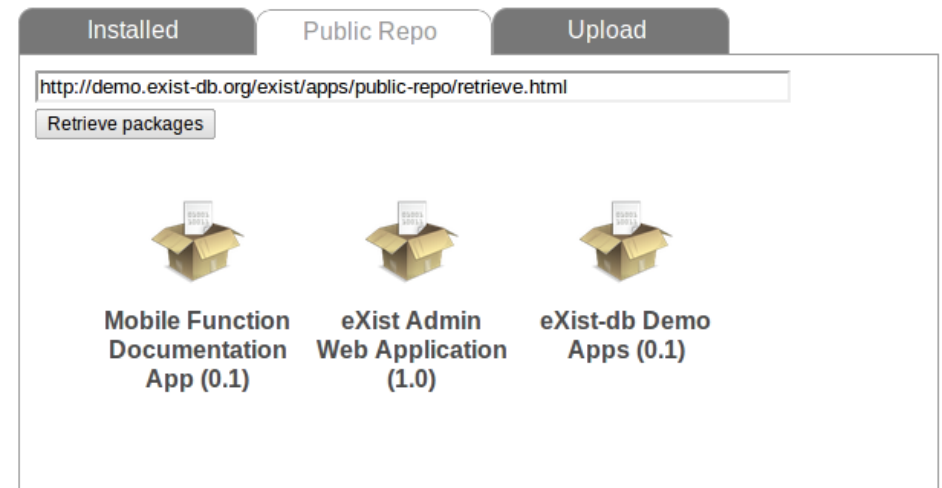
- Modules (libraries)
- Request Processing, URL Rewriting
- MVC and Pipelines
- Web Frameworks
- Deployment options
- Serialization (to XML, HTML, JSON, ...)
- XForms integration

Application Repository



- Self-contained, modular applications
- Install an app with a click
- Everything lives inside the database

Package Repository



Application Repository



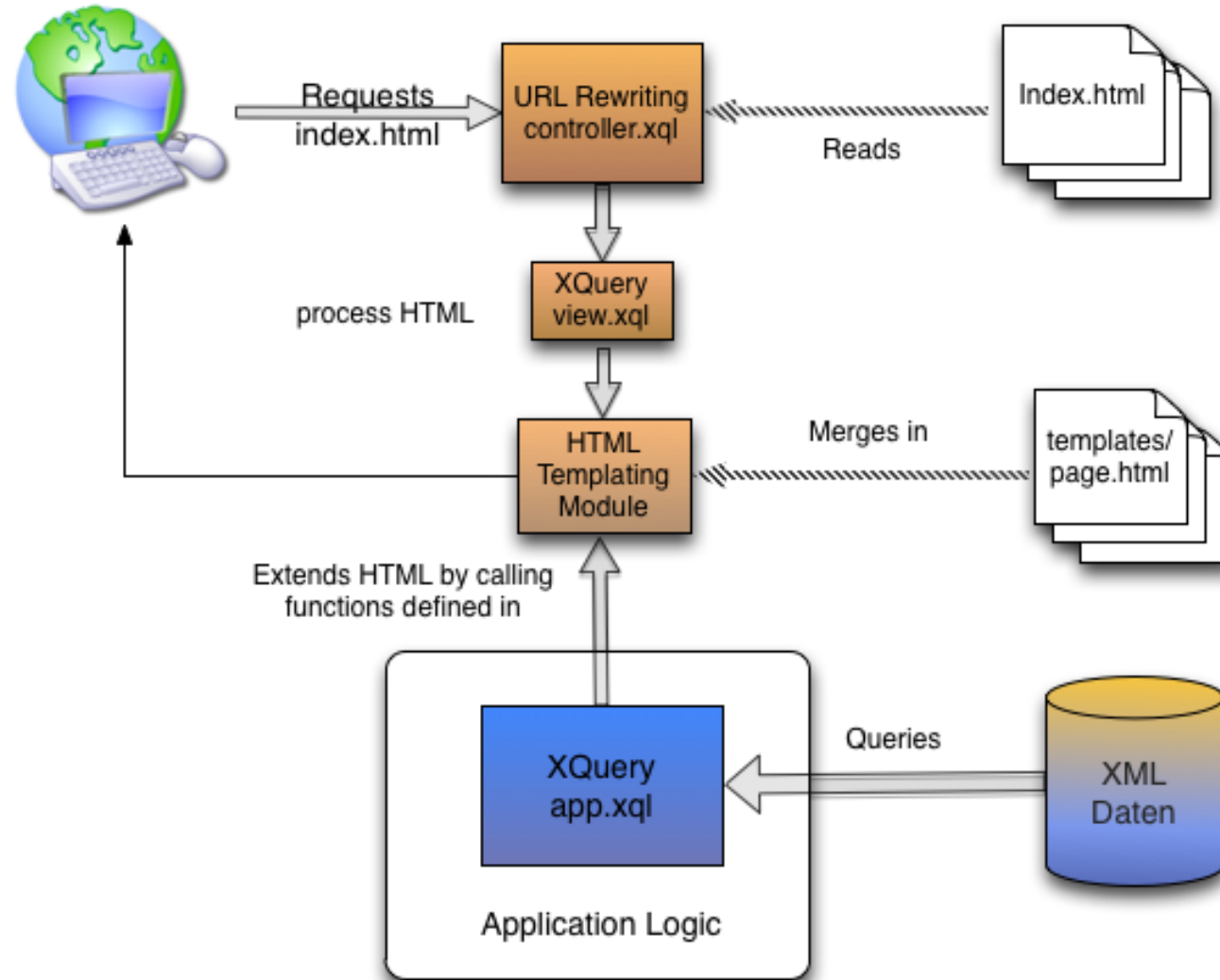
- Make it easy for users to create new apps

The screenshot shows a 'Deployment Editor' window with the following fields and values:

- Template:** HTML Templates (dropdown)
- Source Collection:** /db/myapp
- Target collection:** /db/myapp (with a green checkmark icon)
- Name:** http://exist-db.org/myapp (with a green checkmark icon)
- Abbreviation:** myapp
- Title:** MyApp (with a green checkmark icon)
- Version:** 0.1 (with a green checkmark icon)
- Status:** Alpha (dropdown)
- Pre-install XQuery:** pre-install.xql (with a green checkmark icon)
- Post-install XQuery:** post-install.xql (with a green checkmark icon)

At the bottom right, there are four buttons: Back, Next, Cancel, and Done.

Example Request Pipeline





- Separation of concerns: do not mix HTML pages with application logic (in XQuery)
- Model View Controller (MVC) concept
 - M: the XML data
 - V: the HTML page(s)
 - C: XQuery modules implementing application logic

Quick Example



```
1
2 <div class="templates:surround?with=templates/page.html&at=main">
3   <div id="content">
4     <div class="app:work">
5       <h1>
6         <a href="view-play.html" class="app:work-title"/>
7       </h1>
8       <div class="app:outline"/>
9     </div>
10  </div>
11  <div id="right">
12    <form id="quicksearch" method="POST" action="search.html">
13      <fieldset>
14        <input type="search" name="query" placeholder="Search string" class
15        |="templates:form-control"/>
16      </fieldset>
17      <a href="search.html">Advanced search</a>
18    </form>
19  </div>
20</div>
```

HTML Templating



```
<div class="templates:surround?with=templates/page.html&at=content">
```

calls



```
declare function templates:surround($node as node(),  
    $params as element(parameters)?, $model as item(*)
```

- Templating searches known modules for matching function
- Function signature is the same for all template functions



So, XRX is cool, but we're not
interested in web apps!



XRX Beyond the Web



Case Study 1: Generating an ebook on the fly

Shakespeare Plays in TEI



- Web application to demonstrate query processing and transformations

The screenshot shows the eXistdb web application interface for Shakespeare's Plays. The header includes the eXistdb logo and the text "Open Source Native XML Database". Below the header, there is a section titled "Shakespeare's Plays" with a "List of Works" table. The table has three columns: ID, Title, and Download. The Download column contains links to download each play as an EPUB file. On the right side of the interface, there is a search bar with a "Search string" input field and a "Advanced search" button. Below the search bar, there is a disclaimer box stating that the application is based on data provided by The WordHoard Shakespeare, a joint project of the Perseus Project at Tufts University, The Northwestern University Library, and Northwestern University Academic Technologies. The Northwestern University logo is also present in the disclaimer box.

ID	Title	Download
sha-lov	A Lover's Complaint	epub
sha-mnd	A Midsummer Night's Dream	epub
sha-aww	All's Well That Ends Well	epub
sha-ant	Antony and Cleopatra	epub
sha-ayl	As You Like It	epub
sha-cor	Coriolanus	epub
sha-cym	Cymbeline	epub
sha-ham	Hamlet, Prince of Denmark	epub
sha-juc	Julius Caesar	epub
sha-kil	King Lear	epub
sha-lll	Love's Labour's Lost	epub
sha-mac	Macbeth	epub
sha-mem	Measure for Measure	epub
sha-man	Much Ado About Nothing	epub
sha-oth	Othello, the Moor of Venice	epub
sha-per	Pericles, Prince of Tyre	epub
sha-roj	Romeo and Juliet	epub
sha-son	Sonnets	epub
sha-com	The Comedy of Errors	epub

Exercise



Extend the demo so a given play can be retrieved in two different formats using REST URIs:

- Retrieve HTML:

/shakes/plays/sha-mac.html

- Retrieve epub:

/shakes/plays/sha-mac.epub

Structure of an epub



```
Archive: sha-ham.epub
  Length      Date    Time    Name
-----
      20  2012-03-03 12:12  mimetype
     259  2012-03-03 12:12  META-INF/container.xml
    1538  2012-03-03 12:12  OEBPS/content.opf
    2670  2012-03-03 12:12  OEBPS/title.html
     903  2012-03-03 12:12  OEBPS/table-of-contents.html
   128953  2012-03-03 12:12  OEBPS/sha-ham1.html
   108906  2012-03-03 12:12  OEBPS/sha-ham2.html
   141496  2012-03-03 12:12  OEBPS/sha-ham3.html
    98394  2012-03-03 12:12  OEBPS/sha-ham4.html
   120409  2012-03-03 12:12  OEBPS/sha-ham5.html
    1149  2012-03-03 12:12  OEBPS/styleSheet.css
    7101  2012-03-03 12:12  OEBPS/toc.ncx
-----
    611798
                                12 files
```

Implementation



- For generating the HTML, simply reuse the transformation we wrote for the web view
- Use `compression:zip` to pack the files
- Only two additional XQuery files needed:

<code>epub.xql</code>	Library module to generate all utility files needed for the epub
<code>get-epub.xql</code>	Service which returns a complete epub for a given work id.

Implementation



- The only complex function is the one to generate the table of contents:

```
244 declare function epub:toc-ncx-div($root as element(), $start as xs:int) {
245     for $div in $root/tei:div
246     let $id := $div/@xml:id
247     let $file := $div/ancestor-or-self::tei:div[last()]/@xml:id
248     let $index := count($div/preceding::tei:div[ancestor::tei:body]) + count($div/ancestor::tei:div) + 1
249     return
250         <navPoint id="navpoint-{$id}" playOrder="{ $start + $index}" xmlns="http://www.daisy.org/z3986/2005/ncx/">
251             <navLabel>
252                 <text>{$div/tei:head/text()}</text>
253             </navLabel>
254             <content src="{ $file }.html#{ $id }"/>
255                 { epub:toc-ncx-div($div, $start) }
256             </navPoint>
257 };
258
```

Summary



- Generating the epub from the existing web site does not require any advanced XQuery skills
- Most of the code can be reused
- Fast enough to do the transformation on the fly
- EPUB 2.0 limits your possibilities though



3. Memorandum for the President's Files by the President's Assistant for National Security Affairs (Kissinger)¹

Washington, January 30, 1973, 9:30–10 a.m.

SUBJECT

Meeting with South Vietnamese Foreign Minister Tran Van Lam, Tuesday, January 30, 1973, at 9:30–10:00 a.m.

PARTICIPANTS

The President
Dr. Henry A. Kissinger
Foreign Minister Tran Van Lam
Ambassador Tran Kim Phuong

[After a brief photo opportunity, the President greeted Foreign Minister Lam, who was on his way back to Saigon after signing the Ceasefire Agreement in Paris on January 27.]²

The President: We must have an even closer relationship as peacetime allies. The settlement will last only as long as our two governments go forward together. You can count on our continued military and material support and economic support. And spiritual support. We recognize only one government in South Vietnam. The Republic of Vietnam will be recognized and assisted. The key to U.S. policy in Indochina is our continued alliance and friendship with Vietnam. I have sent the Vice President to show that we are standing firm with our allies.³ You have stood firm. We respect you for it. The American press shouldn't discourage you. They don't represent the American people. You should know you have a friend in this office.

We all have a responsibility to cool it now, however, China and the USSR will be urged to restrain their friends.

Foreign Minister Lam: We are gratified by your approach. Sometimes we gave a hard time to Dr. Kissinger. But we perfectly understand the necessity of sticking together. We had to show we tried to get the maximum we could. But it finally depends on good will between us.

62 / 1372

- Original code used at *US State Department* to publish the *Foreign Relations of the United States* series in e-book format

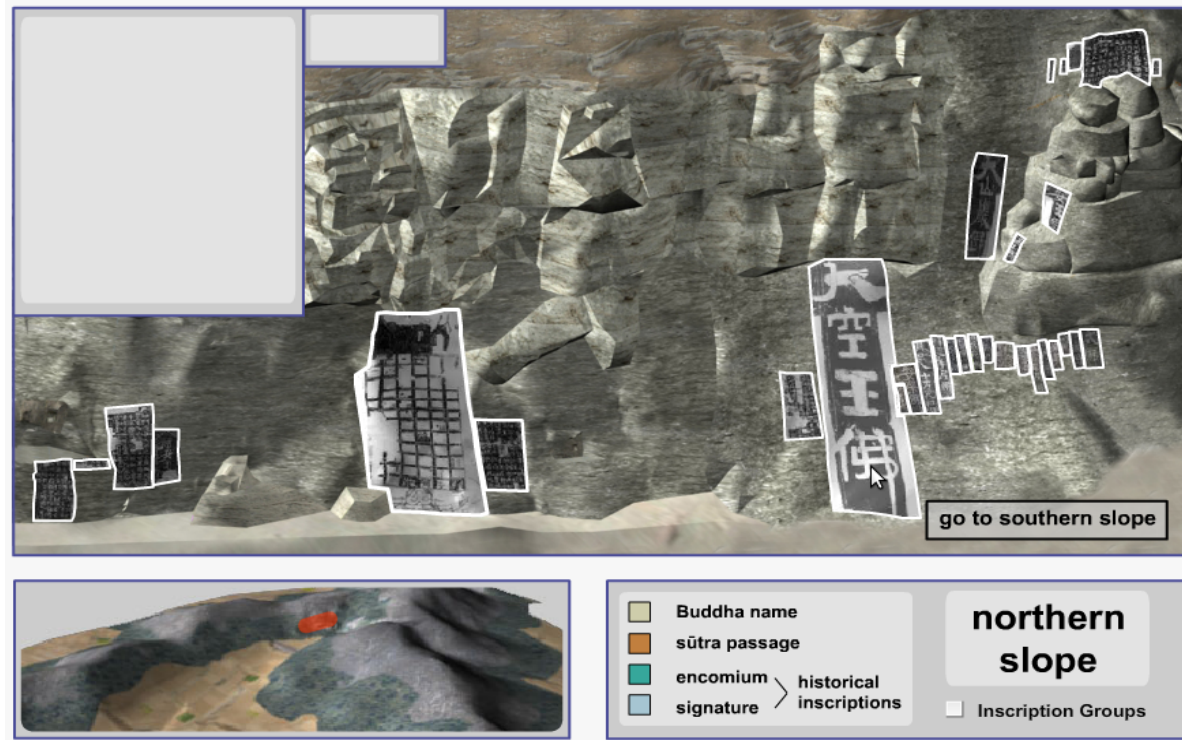


Case Study 2: Buddhist Stone Sutras in China

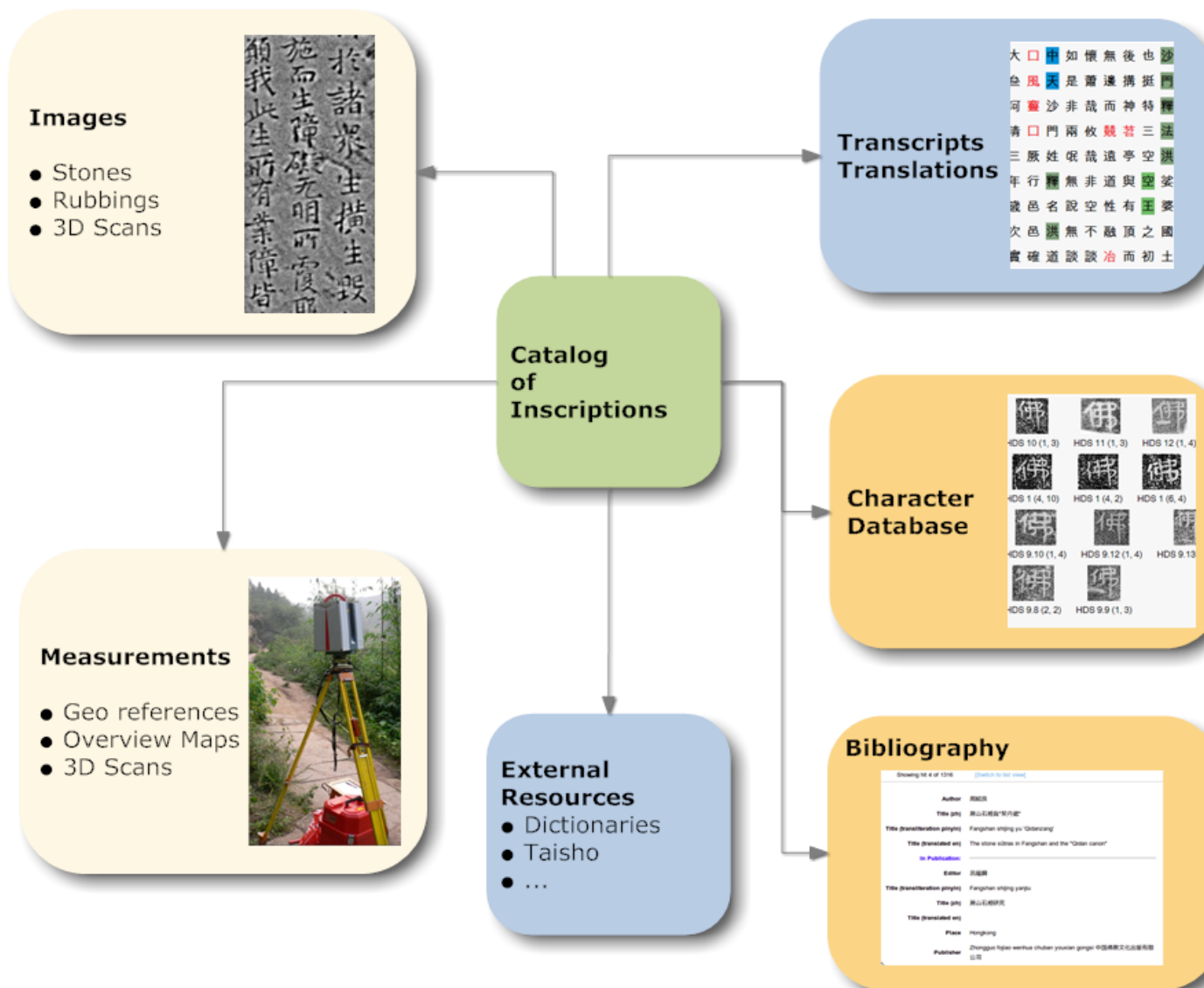
Buddhist Stone Sutras in China



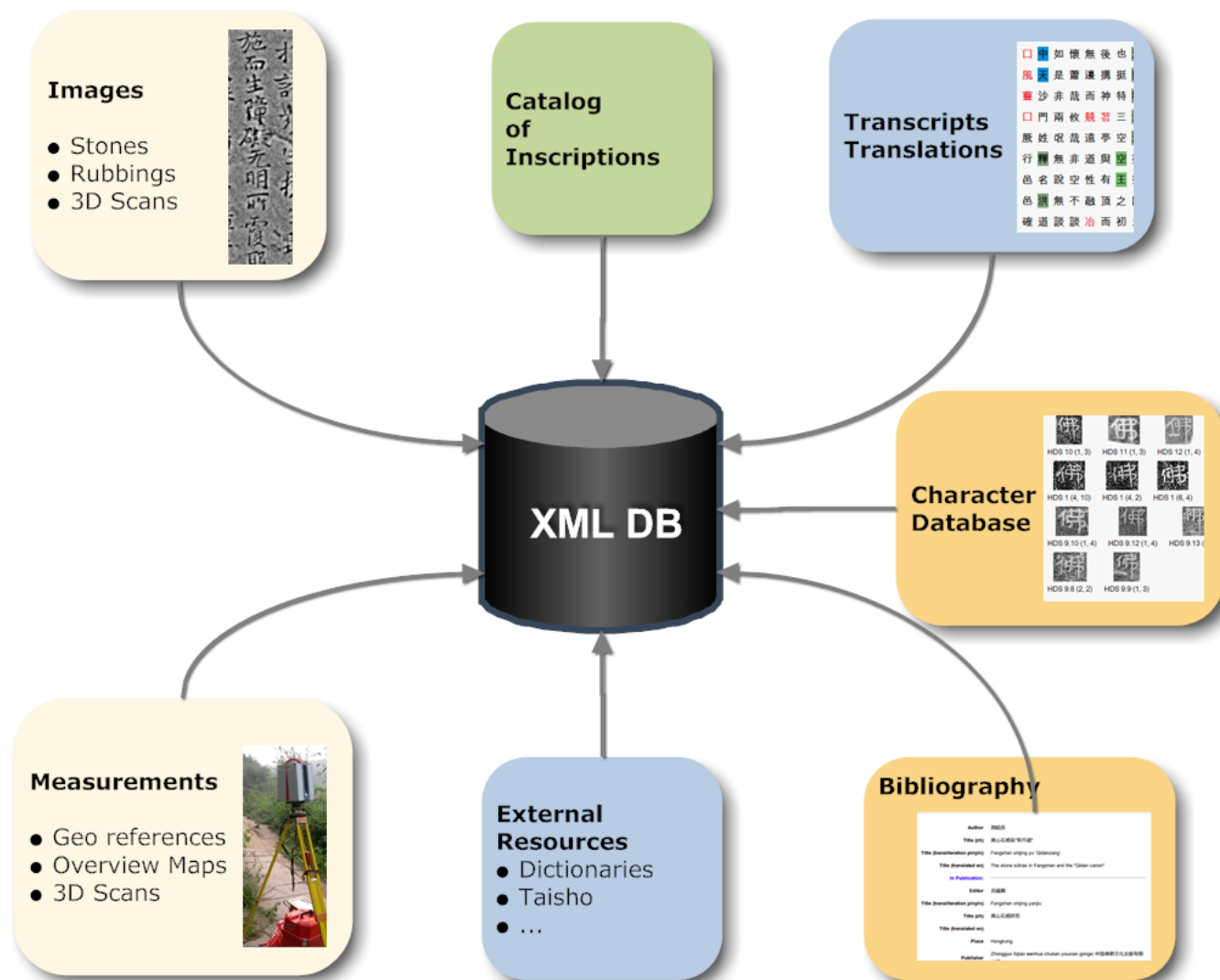
- Heidelberger Akademie der Wissenschaften
- Documentation of Buddhist stone inscriptions in China



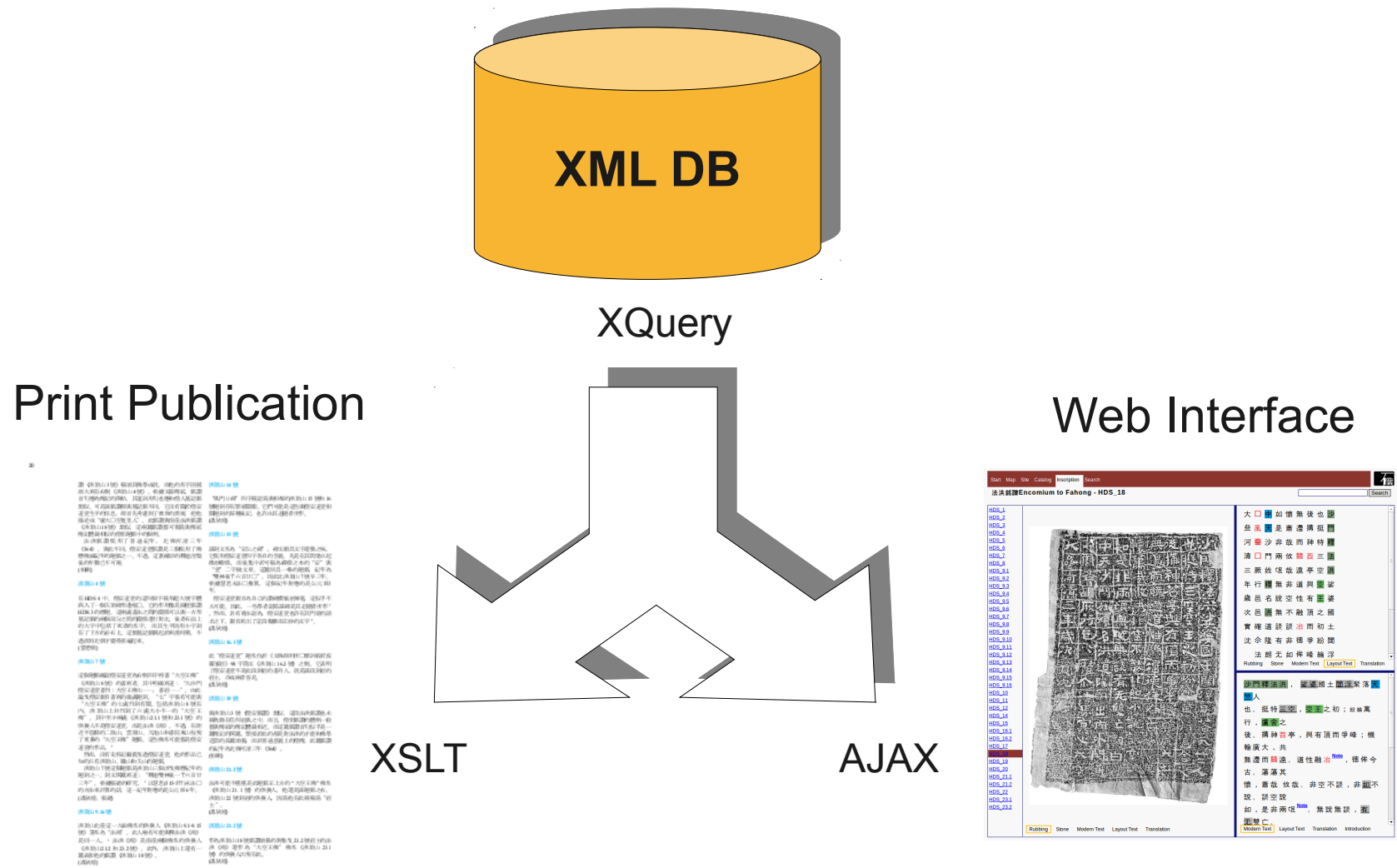
Data Collected



One Database for Everything



Cross-Media Publishing



Publication Goals

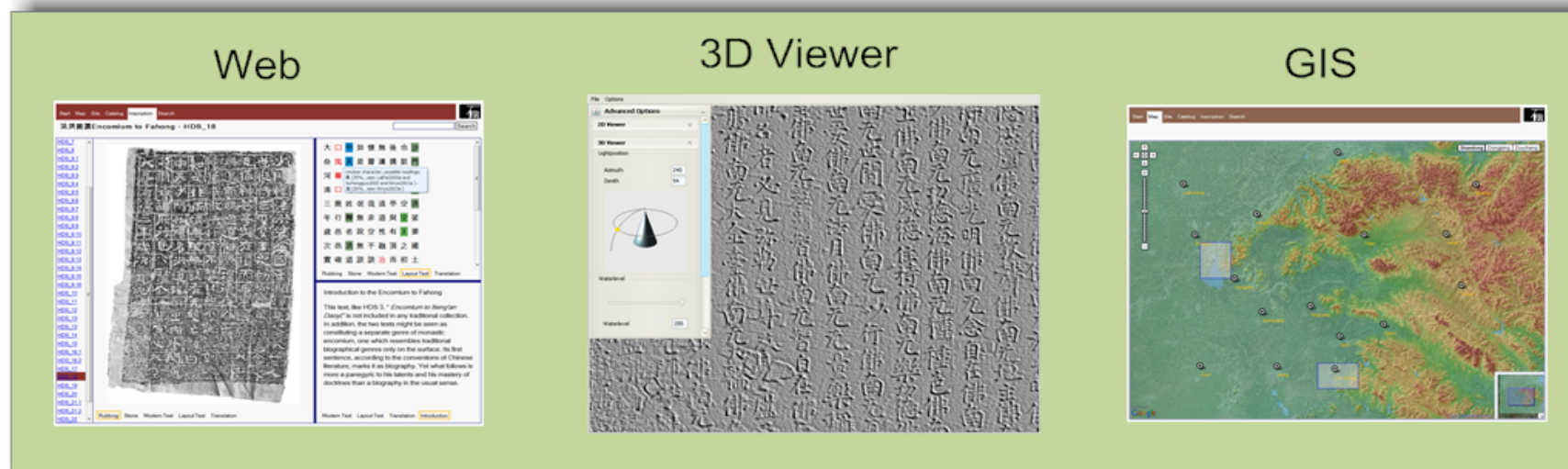


- Print publication presents final research results
- Web application: access to all materials which led to those results

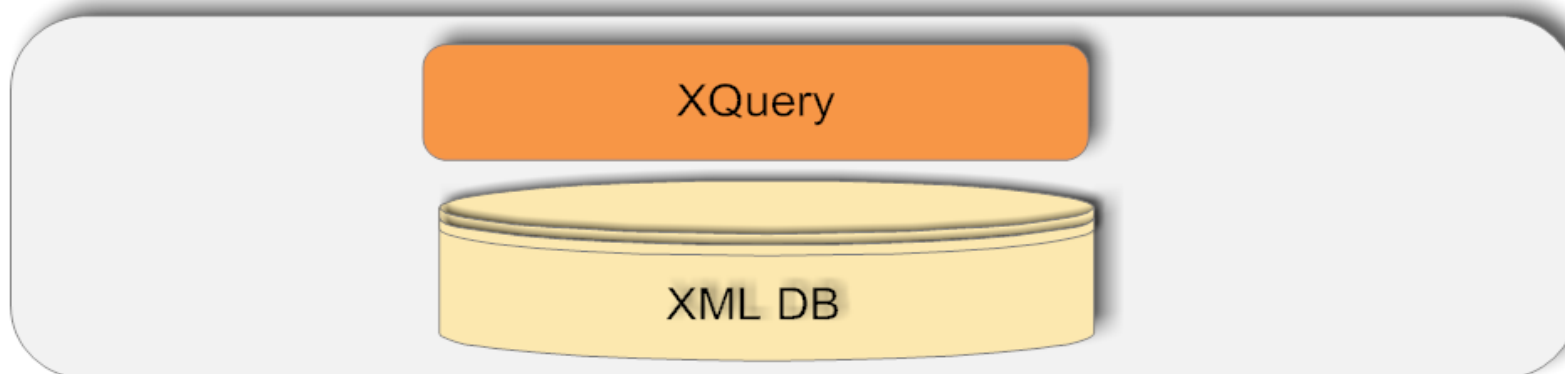


- Enable users to follow and verify every transcription and interpretation
- A virtual experience allowing users to explore inscriptions within their environment
 - 2D GIS showing inscriptions in context
 - 3D models of sites/caves
 - Interactive viewer for scans
 - Compare scans, transcriptions and translations

XRX in the Web App



Services



The TEI Markup



- Nearly every character is annotated
- Unclear characters, different readings
- Dates, names and titles of persons and places ...

```
<lb n="1"/>
<persName corresp="en-a1" key="rh0004" type="monk" xml:id="zh-a1">
  <addName type="honorific" rend="up">大</addName>
  <roleName type="monk">沙門</roleName>
  <addName corresp="en-a2" type="clerical" xml:id="zh-a2">僧安</addName>
</persName> , <choice>
  <unclear cert="0.5" reason="damage" resp="R-Default R2 Ki1 LT LF3 Zhang2">又</unclear>
  <unclear cert="0.3" reason="damage" resp="R3 R4 Ki2 LF1">义</unclear>
  <unclear cert="0.2" reason="damage" resp="YH">氏</unclear>
</choice>
<g rend="variant">名</g>
<persName corresp="en-a3" key="rh0004" type="monk" xml:id="zh-a3">
  <addName corresp="en-a4" type="clerical" xml:id="zh-a4">道壹</addName>
</persName> , <g rend="variant">廣</g>大鄉<choice>
  <unclear cert="0.5" resp="R-Default R3 Zhang2">空</unclear>
  <unclear cert="0.3" resp="LF3 YH R2">□</unclear>
  <unclear cert="0.1" resp="Zhang1 Ki2">佛</unclear>
  <unclear cert="0.1" resp="LT">嶺</unclear>
</choice>
<lb n="2"/>
<space extent="1"/>
```

Challenges for Print



- Static, no interactivity
- Limited cross-referencing
- Thousands of text critical annotations
- No support for traditional Chinese layout
- Huge inscriptions

Challenges for Print



Switching scripts and fonts

the *Treatise on the Great Perfection of Wisdom* (*Da zhidu lun* 大智度論, **Mahā-prajñāpāramitā-sāstra*) that is traditionally attributed to Nāgārjuna (ca. 150 – 250 CE) and was supposedly translated by Kumārajīva.⁷¹ In the Dunhuang manuscripts, this scripture's head title is *Da zhidu lun*, while its end title is *Mahāyāna-sūtra*.⁷² A passage quoted and commented upon in

Reproducing vertical text layout

亦	亦	亦	國	衆	羅	佛
名	名	名	玉	生	蜜	告
龍	護	辟	之	心	是	波
寶	國	鬼	父	識	諸	斯
神	珠	珠	母	之	佛	匿
玉	亦	亦	也	神	菩	玉
	名	名	亦	本	薩	是
	天	如	名	也	一	般
	地	意	神	一	切	若
	鏡	珠	符	切		波

Challenges for Print



All information has to be Chinese/English:

大正藏 Taisho

T #1332, 21, 《大正藏》21冊, 1332號, 第536頁c3 21:536c3

T #1336, 21, 《大正藏》21冊, 1336號, 第627頁a20 21:627a20 《大正藏》21冊, 1336號, 第627頁c28 21:627c28

T #1339, 21, 《大正藏》21冊, 1339號, 第650頁c3 21:650c3 《大正藏》21冊, 1339號, 第653頁a3-13 21:653a3-13

T #1393, 21, 《大正藏》21冊, 1393號, 第910頁b16 21:910b16

T_2045, 《大正藏》50冊, 2045號, 第178頁a13 50:178a13

錄文 Transcriptions

ZHANG ZONG 1997, 113

圖示 Illustrations

LIU WENJIN UNPUBLISHED, 10

LIU WENJIN 1997, 26-27

討論 Discussion

AN TINGSHAN 1997, 128

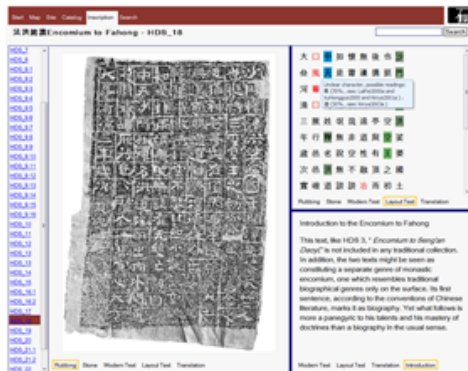
安廷山簡略提到了該銘文。

An Tingshan briefly mentions this inscription.

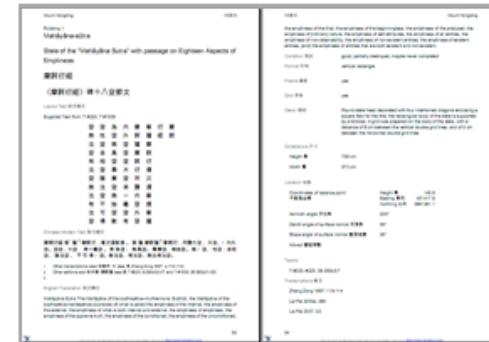
Implementation



Web Application



Print



Services

PDF Export

XQuery

XSL FO

XML DB



- Various XQuery services deliver content chunks to the client using HTTP/REST
- Client:
 - Web page: integrate content using AJAX
 - Print: XSLT stylesheets generate XSL FO

Example: Layout Text



- Writing direction: top-right to bottom-left
- Direction sometimes changes to fit space
- Neither properly supported by HTML nor XSL FO



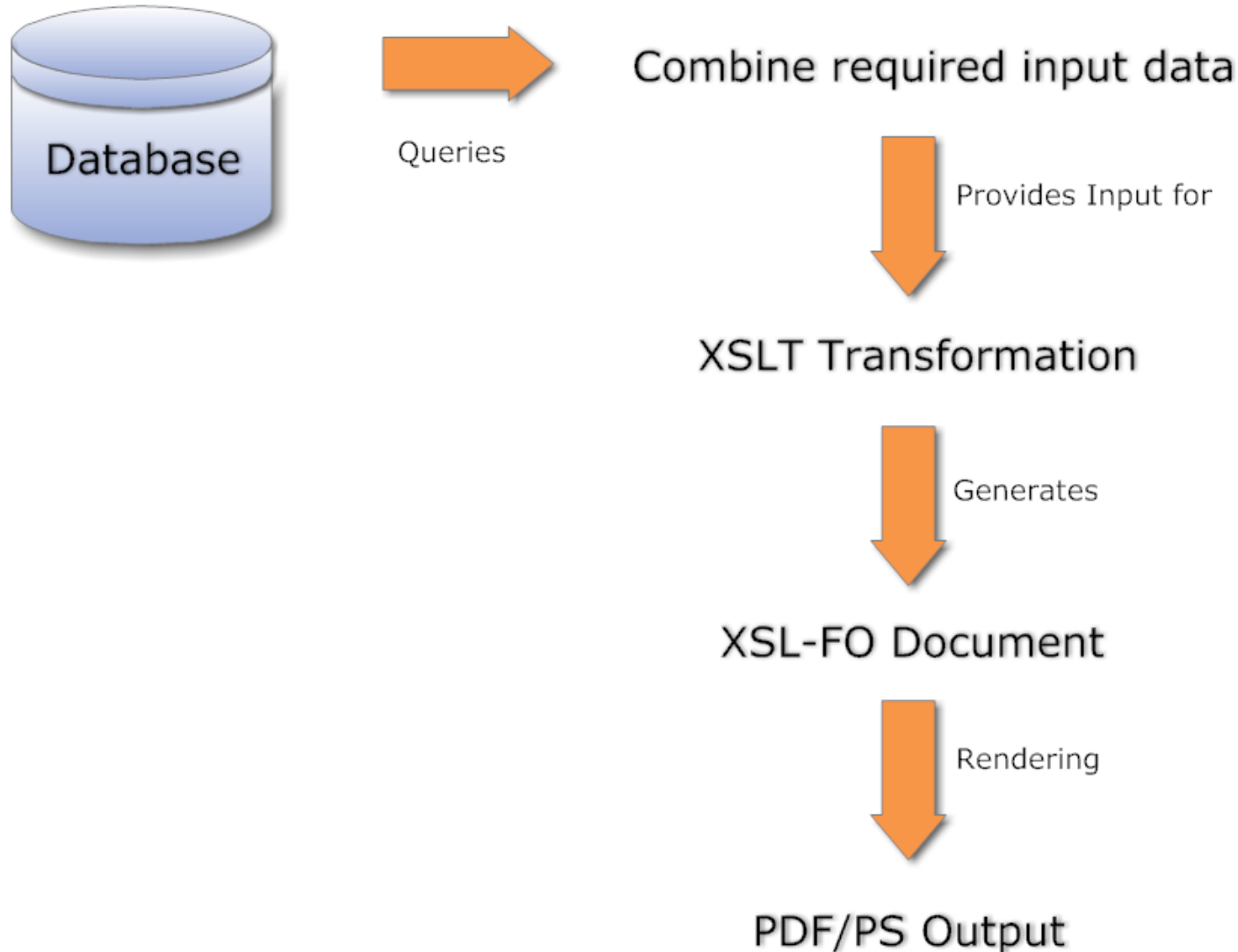
Solution



- Use an XQuery to render established reading in proper layout
- Generates SVG output
- Directly integrated into XSL FO

大 口 中 如 懷 無 後 也 沙
叁 風 天 是 蕭 邊 構 挺 門
河 壺 沙 非 哉 而 神 特 釋
清 口 門 兩 攸 競 苕 三 法
三 厥 姓 氓 哉 遠 亭 空 洪
年 行 釋 無 非 道 與 空 婆
歲 邑 名 說 空 性 有 王 婆
次 邑 洪 無 不 融 頂 之 國
實 確 道 談 談 冶 而 初 土
沈 尔 隆 有 非 德 爭 紛 閭

The Publication Pipeline



Results



- Update all publication targets at the press of a button!
- Changes are immediately propagated to print and web
- Researchers concentrate on their data and texts, not presentation
- Easy integration of other data sources
- Support exchange between projects



Thank you!

<http://exist-db.org>

<http://dev.stonesutras.org>