

Data Repositories in a Virtual Research Environment

- The future of New Women Writers -

- Ronald Haentjens Dekker
- Gertjan Filarski

Development Team - Huygens Institute for the History of the Netherlands

13-04-2011



Content

Introduction

- Who we are
- History of the current application

Virtual Research Environment

- User requirements
- Virtual Research Environment
- Scientific benefits, costs & efficiency

Implementation

- Metadata

Questions & Discussion



Introduction – Who we are

The Huygens Institute conducts research in the fields of Dutch letters and the history of Dutch science.

Institute for the History of the Netherlands engaged in research into the history of the Netherlands.

January 1st merged into **The Huygens Institute for the History of the Netherlands**
Part of the **KNAW** – Royal Netherlands Academy of Arts and Sciences

Largest institute for humanities in The Netherlands:

- About 100 people: 85 scholars and 10 software developers
- Development team works on 5 to 6 projects at any given time.



Introduction – History of the current application

New Women Writers:

- Design and development by Ben Brandenburg, University of Utrecht
- Hosting and maintenance transferred to Huygens ING in 2010
- Lightweight Ruby On Rails application

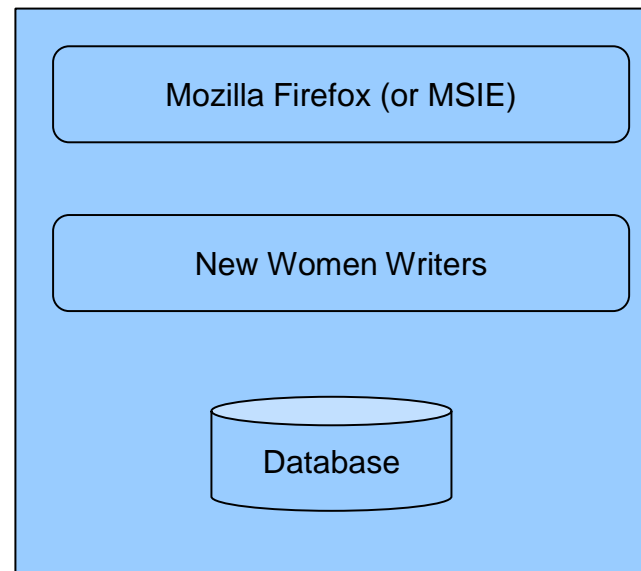
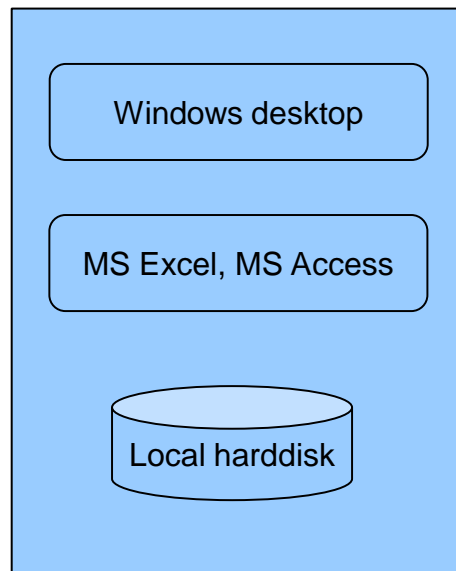
- No serious issues or bugs
- Known time-outs with large numbers of Receptions and large Excel exports
- Known to generate 'empty' records

- Small number of requested new features



Introduction – Premise

Two common user scenario's:



Questions from researchers:

How do I find research data?

How do I check its history?

How do I collaborate with others on a research project?

How do I use data stored in another system for my own research?

Challenge: make research data usable outside its original application

Virtual Research Environment – User requirements

Discoverability: making datasets discoverable by adding interpretable metadata to increase reuse and decrease the chance of developing services that overlap in function.

Persistency: unique and permanent identification of data.

Accessibility: 'read' user access to the full scope of interconnected data sets through a single sign-on in a virtual research environment.

Versioning: management of changes to data over time.

Stability: reliable long term storage for: sources, (heterogeneous) data and standardized metadata.

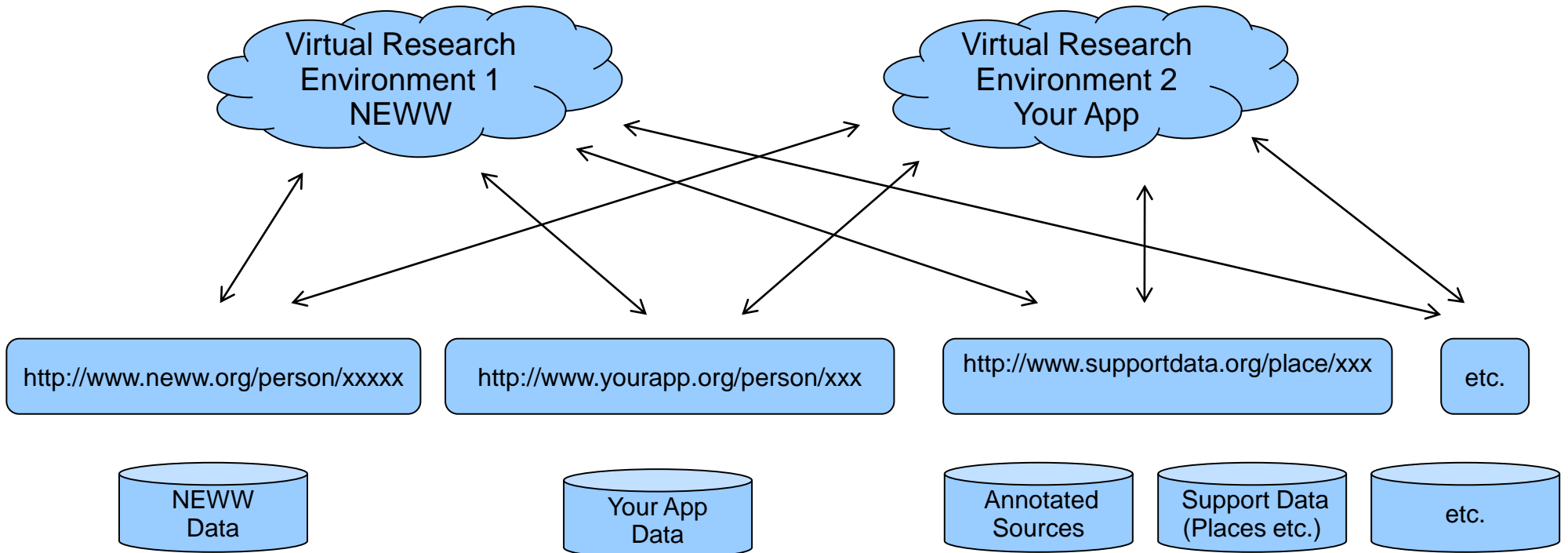
Archiving: a reliable and qualified archiving service for storage in a platform independent format.

Extendibility: easy and low cost accommodation of new data sets in existing or new repositories.

By accomplishing these goals Huygens ING meets all requirements for its status as CLARIN center.



Virtual Research Environment



Data Repository: a logical partitioning of data from multiple sources which are persistently accessible through the World Wide Web.

Virtual Research Environment



Where were they
at what time?

Persons repository

Biography

List of publications



Virtual Research Environment - **Scientific benefits**

Approach: different communities of scholars ask different questions to the same sets of data.

Access: independent storage of (heterogeneous) data avoids technical limitations of research.

Shared Tools: analytical tools are shared among research communities.

Focus: projects / communities are only responsible for the quality of data in their field. Supporting data is provided by others.

Scope: connected sets of data allow for better understanding and improved vision in a wider context, increasing the chance of discovery.



Implementation - Metadata

Metadata: data providing information about one or more aspects of data.

The Simple **Dublin Core Metadata Element Set (DCMES):**

<i>Title</i>	<i>Creator</i>	<i>Subject</i>
<i>Description</i>	<i>Publisher</i>	<i>Contributor</i>
<i>Date</i>	<i>Type</i>	<i>Format</i>
<i>Identifier</i>	<i>Source</i>	<i>Language</i>
<i>Relation</i>	<i>Coverage</i>	<i>Rights</i>

Not enough for a proper definition of e.g. an Author, Work or Reception.



Implementation - Metadata

Metadata: common agreement about the description of an entity

Environment 1:

Author
 Given Name
 Initials
 Family Name
 Birth date
 Birth Place
 Country of Origin
 Bio
 ...

Metadata

Author
 Normalized Name
 Birth date
 Birthplace
 Bio
 ...

Environment 2:

Author
 Given Name
 Family name
 Intraposition
 Birth date
 Birth Place
 Bio
 ...



Questions - Discussion

- Ronald Haentjens Dekker (ronald.dekker@huygensinstituut.knaw.nl)
- Gertjan Filarski (gertjan.filarski@huygensinstituut.knaw.nl)

