Deliverable D5.1
Supplement

HOPE
Grant agreement no: 250549
Heritage of the People’s Europe

Repository Infrastructure and Detailed Design
Appendixes

- Deliverable number: D5.1
- Status: FINAL
- Authors: Jerry de Vries
- Delivery Date: 01-04-2011
- Dissemination level: Public
Version history

<table>
<thead>
<tr>
<th>Date</th>
<th>Changes</th>
<th>Version</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>25-03-2011</td>
<td>-</td>
<td>1.0</td>
<td>Jerry de Vries</td>
</tr>
</tbody>
</table>

Contributors

<table>
<thead>
<tr>
<th>Institution</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>IISG</td>
<td>Gordan Cupac</td>
</tr>
<tr>
<td></td>
<td>Mario Mieldijk</td>
</tr>
<tr>
<td></td>
<td>Sjoerd Siebinga</td>
</tr>
<tr>
<td></td>
<td>Titia van der Werf</td>
</tr>
<tr>
<td></td>
<td>Lucien van Wouw</td>
</tr>
<tr>
<td>CNR-ISTI</td>
<td>Alessia Bardi</td>
</tr>
<tr>
<td></td>
<td>Paolo Manghi</td>
</tr>
<tr>
<td></td>
<td>Franco Zoppi</td>
</tr>
</tbody>
</table>
Table of Contents
Appendix A – Example HOPE Persistent Identifier Web service interface .......... 3
Appendix B – Low Level Design ........................................................................................................... 8
Appendix C – Organizations providing parts of the infrastructure of the SOR .... 16
Appendix D – Technical Glossary SOR .................................................................................................. 17
Appendix A – Example HOPE Persistent Identifier Web service interface

```xml
<?xml version="1.0" encoding="UTF-8" standalone="no" ?>
<wsdl:definitions xmlns:wsdl="http://schemas.xmlsoap.org/wsdl/

xmlns:sch="http://www.iisg.org/schemas/handle"
xmlns:soap="http://schemas.xmlsoap.org/wsdl/soap/"
xmlns:tns="http://www.iisg.org/schemas/handle"
targetNamespace="http://www.iisg.org/schemas/handle">
  <wsdl:types>
    <xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema"

attributeFormDefault="unqualified" elementFormDefault="qualified"
targetNamespace="http://www.iisg.org/schemas/handle">
    <!-- Elements -->
    <xs:element name="CreatePidRequest" type="tns:CreatePidRequestType" />
    <xs:element name="CreatePidResponse" type="tns:CreatePidResponseType" />
    <xs:element name="GetPidByAttributeRequest" type="tns:GetPidByAttributeRequestType" />
    <xs:element name="GetPidByAttributeResponse" type="tns:GetPidByAttributeResponseType" />
    <xs:element name="GetPidRequest" type="tns:GetPidRequestType" />
    <xs:element name="GetPidResponse" type="tns:GetPidResponseType" />
    <xs:element name="UpdatePidRequest" type="tns:UpdatePidRequestType" />
    <xs:element name="UpdatePidResponse" type="tns:UpdatePidResponseType" />
    <xs:element name="GetHopePidRequest" type="tns:GetHopePidRequestType" />
    <xs:element name="GetHopePidResponse" type="tns:GetHopePidResponseType" />
    <!-- Complex Types -->
    <xs:complexType name="locationType">
      <xs:simpleContent>
        <xs:extension base="xs:string">
          <xs:attribute name="href" type="xs:anyURI" use="required" />
          <xs:attribute name="id" type="xs:string" use="optional" />
          <xs:attribute name="weight" type="xs:string" use="optional" />
          <xs:attribute name="view" type="xs:string" use="optional" />
          <xs:attribute name="country" type="xs:string" use="optional" />
          <xs:attribute name="weighted" type="xs:string" use="optional" />
        </xs:extension>
      </xs:simpleContent>
    </xs:complexType>
    <xs:complexType name="locAttType">
      <xs:sequence>
        <xs:element maxOccurs="unbounded" minOccurs="0" name="location" type="tns:locationType" />
      </xs:sequence>
    </xs:complexType>
  </xs:schema>
</wsdl:types>
</wsdl:definitions>
```
<xs:complexType name="pidType">
  <xs:sequence>
    <xs:element name="id" type="xs:anyURI" />
    <xs:element name="localIdentifier" type="xs:string" />
    <xs:element name="resolveUrl" type="xs:anyURI" />
    <xs:element name="locAtt" type="tns:locAttType" />
  </xs:sequence>
</xs:complexType>

<xs:complexType name="attributeType">
  <xs:sequence>
    <xs:element name="href" type="xs:string" />
  </xs:sequence>
</xs:complexType>

<xs:complexType name="CreatePidRequestType">
  <xs:sequence>
    <xs:element name="resolveUrl" type="xs:anyURI" />
    <xs:element name="locAtt" type="tns:locAttType" />
  </xs:sequence>
</xs:complexType>

<xs:complexType name="CreatePidResponseType">
  <xs:sequence>
    <xs:element maxOccurs="1" minOccurs="0" name="pid" type="tns:pidType" />
  </xs:sequence>
</xs:complexType>

<xs:complexType name="GetPidByAttributeRequestType">
  <xs:sequence>
    <xs:element name="attribute" type="tns:attributeType" />
  </xs:sequence>
</xs:complexType>

<xs:complexType name="GetPidByAttributeResponseType">
  <xs:sequence>
    <xs:element maxOccurs="1" minOccurs="0" name="pid" type="tns:pidType" />
  </xs:sequence>
</xs:complexType>

<xs:complexType name="GetPidRequestType">
  <xs:sequence>
    <xs:element name="id" type="xs:anyURI" />
  </xs:sequence>
</xs:complexType>

<xs:complexType name="GetPidResponseType">
  <xs:sequence>
    <xs:element maxOccurs="1" minOccurs="0" name="pid" type="tns:pidType" />
  </xs:sequence>
</xs:complexType>

<xs:complexType name="UpdatePidRequestType">
  <xs:sequence>
    <xs:element name="pid" type="tns:pidType" />
  </xs:sequence>
</xs:complexType>
HOPE is co-funded by the European Union through the ICT Policy Support Programme.
HOPE is co-funded by the European Union through the ICT Policy Support Programme.
<wsdl:input name="UpdatePidRequest">
  <soap:body use="literal" />
</wsdl:input>

<wsdl:output name="UpdatePidResponse">
  <soap:body use="literal" />
</wsdl:output>

<wsdl:operation name="GetPidByAttribute">
  <soap:operation soapAction="" />
  <wsdl:input name="GetPidByAttributeRequest">
    <soap:body use="literal" />
  </wsdl:input>
  <wsdl:output name="GetPidByAttributeResponse">
    <soap:body use="literal" />
  </wsdl:output>
</wsdl:operation>

<wsdl:operation name="GetPid">
  <soap:operation soapAction="" />
  <wsdl:input name="GetPidRequest">
    <soap:body use="literal" />
  </wsdl:input>
  <wsdl:output name="GetPidResponse">
    <soap:body use="literal" />
  </wsdl:output>
</wsdl:operation>

<wsdl:service name="HandleResourceService">
  <wsdl:port binding="tns:HandleResourceSoap11" name="HandleResourceSoap11">
    <soap:address location="http://195.169.122.195:80/pidservice/" />
  </wsdl:port>
</wsdl:service>
</wsdl:definitions>
Appendix B – Low Level Design

Low level design overview
Low level design specified
Virtual Environment Management

Datacenter IISG Amsterdam

SOR VM environment Management

STORAGEnet

Datacenter Vancis Amsterdam

SOR VM environment Management

STORAGEnet
VM environment applications datacenter IISG
VM environment application datacenter Vancis

HOPE is co-funded by the European Union through the ICT Policy Support Programme.
VM environment converter

[Schematic diagram of VM environment converter with labels for Converter Manager Images and Converter Manager Movies]
VM storage environment

Datacenter IISG Amsterdam

Datacenter Vancis Amsterdam

HOPE is co-funded by the European Union through the ICT Policy Support Programme.
Network overview

HOPE is co-funded by the European Union through the ICT Policy Support Programme.
Appendix C – Organizations providing parts of the infrastructure of the SOR

IISG

International Institute of Social History is the institute that collects and preserves collections and material related to cultural heritage and supports the effort to safeguard an international cultural heritage that still is, too often, in danger of disappearing.

The IISG provides the developers to build and implement the SOR and is responsible for the management and development of the software and PID service.

KNAW

Royal Dutch academy of Science promotes quality in science and scholarship and strives to ensure that Dutch scholars and scientists contribute to cultural, social and economic progress. As a research organization, the Academy is responsible for a group of outstanding national research institutes. It promotes innovation and knowledge valorization within these institutes and encourages them to cooperate with one another and with university research groups.

The KNAW manages the service contracts of SURFnet and Vancis datacenter services.

SURFnet

SURFnet provides a secure internet connection and network for researchers, teachers and students to help them cooperate easily and fast. SURFnet provides internet and network facilities (wired and wireless), domain and DNS resolving, IPv6.

The KNAW is connected to SURFnet. SURFnet is the provider of the internet connection of the KNAW and its institutes.

Vancis

Part of the SARA network and data center and provides ict-solutions and products as data back-up and archiving, hosting, data storage, wide area network solutions, virtualization.

Vancis is responsible for providing the virtualization of the SOR components.
## Appendix D – Technical Glossary SOR

<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
<th>Source</th>
<th>See also</th>
</tr>
</thead>
<tbody>
<tr>
<td>Access Rights</td>
<td>The information that identifies the legal access restrictions pertaining to the <strong>HOPE Social History Resource</strong>, relating to legal frameworks such as the Copyright Laws, Privacy Law, etc. and licensing agreements between CPs and rights owners.</td>
<td>Source: HOPE</td>
<td>HOPE Access Conditions Matrix</td>
</tr>
<tr>
<td>Administration Interface</td>
<td>Web interface that provides status information of the SOR to the end user. From this interface a Content Provider is also able to manage and carry out submissions.</td>
<td>Source: HOPE</td>
<td>Administration panel</td>
</tr>
<tr>
<td>Administration Panel</td>
<td>Web interface that provides status information of the SOR to the end user. From this interface a Content Provider is also able to manage and carry out submissions.</td>
<td>Source: HOPE</td>
<td>Administration interface</td>
</tr>
<tr>
<td>API</td>
<td>Application Programming Interface; Set of functions and restrictions that a software program can use and access the service and resources provided by another particular software program that implements that API. It serves as an interface between different software programs</td>
<td>Source: Wikipedia</td>
<td></td>
</tr>
<tr>
<td>Checksum</td>
<td>A numeric representation that is mathematically generated and is used to uniquely identify a collection of data. The md5 algorithm is used for generating the checksum</td>
<td>Source: File Research Center</td>
<td></td>
</tr>
<tr>
<td><strong>Content Provider</strong></td>
<td>A HOPE partner with social history collections which provides metadata and <em>Digital Objects</em> to the <em>HOPE System</em>.</td>
<td>Source: Glossary v2.0</td>
<td></td>
</tr>
<tr>
<td>----------------------</td>
<td>-------------------------------------------------------------------------------------------------</td>
<td>----------------------</td>
<td></td>
</tr>
<tr>
<td><strong>Datacenter</strong></td>
<td>A facility used to house computer systems and associated components, such as telecommunications and storage systems. It generally includes redundant or backup power supplies, redundant data communications connections, environmental controls (e.g., air conditioning, fire suppression) and security devices. Vancis is used as datacenter inside the HOPE project.</td>
<td>Source: Wikipedia</td>
<td></td>
</tr>
<tr>
<td><strong>DBMS</strong></td>
<td>DataBase Management System; A set of computer programs that controls the creation, maintenance, and the use of a database.</td>
<td>Source: Wikipedia</td>
<td></td>
</tr>
</tbody>
</table>
| **Derivative**       | Different versions derived from the Master or from the Born-Digital Original. Derivatives are generally used for web access to digital content. Derivatives may include thumbnail, preview, high- and low-resolution, and OCRed text versions. In HOPE, “Derivative” can be a qualifier; we speak of a “Derivative File” or “Derivative Object” as applicable. | Source: Glossary v2.0
See also: Derivative 1 (2, 3) |
| **Derivative Level 1** | High-resolution Derivative for reproduction and publication (online/print) purposes. | Source: Glossary v2.0
See also: Derivative 1 |
<table>
<thead>
<tr>
<th>Term</th>
<th>Description</th>
<th>Source</th>
<th>See also:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Derivative Level 2</td>
<td>Medium to low-resolution Derivative for online consultation (view/listen) purposes.</td>
<td>Glossary v2.0</td>
<td>Derivative 2</td>
</tr>
<tr>
<td>Derivative Level 3</td>
<td>Preview-quality Derivative (lowest resolution) for display purposes in search results.</td>
<td>Glossary v2.0</td>
<td>Derivative 3</td>
</tr>
<tr>
<td>Derivative Storage</td>
<td>Storage component for derivatives in the HOPE SOR</td>
<td>HOPE</td>
<td>MongoDB</td>
</tr>
<tr>
<td>Dev Lan</td>
<td>Development Local Area Network</td>
<td>HOPE</td>
<td>Local Area Network</td>
</tr>
<tr>
<td>Digital Object Depot</td>
<td>A digital object repository, digital assets management system, or other network accessible system that is used for the ingest, storage, management, and delivery of Digital Objects and that is compliant to a set of agreed minimum functionalities and services within the HOPE system</td>
<td>Glossary v2.0</td>
<td>Digital Object Repository</td>
</tr>
<tr>
<td>Discovery Service</td>
<td>A web portal, which enables the discovery, identification, and selection of materials through searching and browsing functions.</td>
<td>Glossary v2.0</td>
<td></td>
</tr>
<tr>
<td>DMZ</td>
<td>Demilitarized zone is a physical or logical sub network that contains and exposes an organization’s external services to a larger untrusted network, usually the Internet.</td>
<td>Wikipedia</td>
<td></td>
</tr>
<tr>
<td><strong>Drupal</strong></td>
<td>Free and open source content management system (CMS) and Content Management framework (CMF) written in PHP and distributed under the GNU General Public License</td>
<td>Source: Drupal.org</td>
<td></td>
</tr>
<tr>
<td><strong>Event driven architecture</strong></td>
<td>A software architecture pattern promoting the production, detection, consumption of, and reaction to events. This architectural pattern may be applied by the design and implementation of applications and systems which transmit events among loosely coupled software components and services.</td>
<td>Source: Wikipedia</td>
<td></td>
</tr>
<tr>
<td><strong>Firewall</strong></td>
<td>A system designed to prevent unauthorized access to or from a private network. Firewalls can be implemented in both hardware and software, or a combination of both. Firewalls are frequently used to prevent unauthorized Internet users from accessing private networks connected to the Internet, especially intranets. All messages entering or leaving the intranet pass through the firewall, which examines each message and blocks those that do not meet the specified security criteria</td>
<td>Source: Webopedia.com</td>
<td></td>
</tr>
<tr>
<td><strong>Handle</strong></td>
<td>The Handle System provides efficient, extensible, and secure resolution services for unique and persistent identifiers of digital objects. The Handle System includes an open set of protocols, a namespace, and a reference</td>
<td>Source: handle.net</td>
<td></td>
</tr>
</tbody>
</table>
The protocols enable a distributed computer system to store identifiers, known as handles, of arbitrary resources and resolve those handles into the information necessary to locate, access, contact, authenticate, or otherwise make use of the resources. This information can be changed as needed to reflect the current state of the identified resource without changing its identifier, thus allowing the name of the item to persist over changes of location and other related state information.

**HOPE Aggregator**

The system that harvests, stores, and disseminates *Descriptive Metadata* supplied by CPs. The Aggregator enables harmonisation and enrichment of the metadata and provides a *Search API* for use by the *IALHI Portal* and CP institutional websites.

**HOPE Shared Object Repository (SOR)**

The shared *HOPE-Compliant Digital Object Repository* used by some CPs for the ingest, storage, management and delivery of their *Digital Objects*.

**HOPE Staging Area**

A temporarily online storage for Content Providers. A CP can use a SFTP program to upload digital master files to the staging area. On submission the masters will be ingested into the SOR.

**HOPE Technical**

Storage component of the SOR
<table>
<thead>
<tr>
<th>Metadata Storage</th>
<th>where will be stored after ingest</th>
<th>See also: Technical Metadata (Glossary v2.0)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Identification, Authentication, Authorization (IAA)</td>
<td>The SOR has an identification, authentication and authorization system. This is necessary to act on access condition rules, which apply to categories of users in combination with types of usage of digital objects</td>
<td>Source: HOPE</td>
</tr>
<tr>
<td>Ingest</td>
<td>Process whereby, based on the SOR processing instructions, digital objects will be moved from the Staging Area to the SOR Digital Object Repository</td>
<td>Source: HOPE</td>
</tr>
<tr>
<td>ISCSI</td>
<td>Internet Small Computer System Interface; A protocol that serializes SCSI commands and converts them to TCP/IP</td>
<td>Source: Pcmag.com</td>
</tr>
<tr>
<td>Local Area Network (LAN)</td>
<td>A communications network that serves users within a confined geographical area. The &quot;clients&quot; are the user's workstations typically running Windows, although Mac and Linux clients are also used. The &quot;servers&quot; hold programs and data that are shared by the clients. Servers come in a wide range of sizes from Intel-based servers to mainframes.</td>
<td>Source: Pcmag.com</td>
</tr>
<tr>
<td>MongoDB</td>
<td>A scalable, high-performance, open source, document-oriented database.</td>
<td>Source: Mongodb.org</td>
</tr>
<tr>
<td>Term</td>
<td>Definition</td>
<td>Source</td>
</tr>
<tr>
<td>-------------------------------------------</td>
<td>---------------------------------------------------------------------------</td>
<td>-------------------</td>
</tr>
<tr>
<td><strong>Object Oriented Architecture</strong></td>
<td>Computer system design in which all identifiable components (files, operations, processes) may be represented as data structures (objects) in the system’s memory, for manipulation by the system software.</td>
<td>Businessdictionary.com</td>
</tr>
<tr>
<td><strong>Persistent Identifier (PID)</strong></td>
<td>A character string that is globally unique and permanently identifies a resource within a given context. In HOPE, PIDs are always associated with a resolve URL and should be persistently resolvable on the Internet.</td>
<td>Glossary v2.0</td>
</tr>
<tr>
<td><strong>RDBMS</strong></td>
<td>Relational DataBase Management System; A database management system (DBMS) that is based on the relational model as introduced by E. F. Codd. Most popular commercial and open source databases currently in use are based on the relational database model.</td>
<td>Wikipedia</td>
</tr>
<tr>
<td><strong>Router</strong></td>
<td>A device that forwards data packets across computer networks. Routers perform the data “traffic directing&quot; functions on the Internet. A router is a microprocessor-controlled device that is connected to two or more data lines from different networks. When a data packet comes in on one of the lines, the router reads the address information in the packet to determine its ultimate destination</td>
<td>Wikipedia</td>
</tr>
<tr>
<td><strong>SAN</strong></td>
<td>Storage Area Network; A network of storage disks. In large enterprises, a SAN connects multiple servers to a</td>
<td>Pcmag.com</td>
</tr>
</tbody>
</table>
centralized pool of disk storage. Compared to managing hundreds of servers, each with their own disks, SANs improve system administration. By treating all the company’s storage as a single resource, disk maintenance and routine backups are easier to schedule and control. In some SANs, the disks themselves can copy data to other disks for backup without any processing overhead at the host computers.

<p>| Server Lan | A local area network (LAN) server is a program (and by implication usually the computer it runs in) that &quot;serves&quot; the resources (files, storage, application programs, printers, and other devices) for a number of attached workstations. | Source: Wikipedia |
| SQL | Structured Query Language; A database computer language designed for managing data in relational database management systems (RDBMS), and originally based upon relational algebra and calculus | Source: Wikipedia |
| SQL Server | Every DBMS that implements the SQL language | Source: HOPE |
| Switch | A small hardware device that joins multiple computers together within one local area network (LAN). Technically, network switches operate at layer two (Data Link Layer) of the OSI model. | Source: HOPE |</p>
<table>
<thead>
<tr>
<th><strong>Virtualization</strong></th>
<th>Running multiple operating systems on a single machine. While most computers only have one operating system installed, virtualization software allows a computer to run several operating systems at the same time.</th>
<th>Source: techterms.com</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>VLAN</strong></td>
<td>Virtual LAN: A group of PCs, servers and other network resources that behave as if they were connected to a single, network segment, even though they may not be</td>
<td>Source: HOPE</td>
</tr>
<tr>
<td><strong>WAN</strong></td>
<td>Wide Area Network; A long-distance communications network that covers a wide geographic area, such as a state or country. The telephone companies and cellular carriers deploy WANs to service large regional areas or the entire nation. Large enterprises have their own private WANs to link remote offices, or they use the Internet for connectivity.</td>
<td>Source: HOPE</td>
</tr>
<tr>
<td><strong>Webservice</strong></td>
<td>A software system designed to support interoperable machine-to-machine interaction over a network. It has an interface described in a machine-processable format (specifically Web Services Description Language WSDL). Other systems interact with the Web service in a manner prescribed by its description using SOAP messages, typically conveyed using HTTP with an XML serialization in conjunction with other Web-related</td>
<td>Source: W3C</td>
</tr>
</tbody>
</table>

See also: WSDL
| WSDL | WebServices Description Language; The WSDL defines services as collections of network endpoints, or ports. The WSDL specification provides an XML format for documents for this purpose. The abstract definitions of ports and messages are separated from their concrete use or instance, allowing the reuse of these definitions. A port is defined by associating a network address with a reusable binding, and a collection of ports defines a service | Source: W3C  
See also Webservice |