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

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

Version history

Date	Changes	Version	Name
25-03-2011	-	1.0	Jerry de Vries

Contributors

Institution	Name
IISG	Gordan Cupac
	Mario Mieldijk
	Sjoerd Siebinga
	Titia van der Werf
	Lucien van Wouw
CNR-ISTI	Alessia Bardi
	Paolo Manghi
	Franco Zoppi

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

Table of Contents

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

Appendix A – Example HOPE Persistent Identifier Web service interface

```
<?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: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>
```

```
- <xs:complexType name="UpdatePidResponseType">
- <xs:sequence>
  <xs:element maxOccurs="1" minOccurs="0" name="pid" type="tns:pidType" />
    </xs:sequence>
    </xs:complexType>
- <xs:complexType name="GetHopePidRequestType">
- <xs:sequence>
  <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="GetHopePidResponseType">
- <xs:sequence>
  <xs:element maxOccurs="1" minOccurs="0" name="pid" type="tns:pidType" />
    </xs:sequence>
    </xs:complexType>
    </xs:schema>
    </wsdl:types>
- <wsdl:message name="CreatePidRequest">
  <wsdl:part element="tns:CreatePidRequest" name="CreatePidRequest" />
    </wsdl:message>
- <wsdl:message name="GetHopePidRequest">
  <wsdl:part element="tns:GetHopePidRequest" name="GetHopePidRequest" />
    </wsdl:message>
- <wsdl:message name="UpdatePidRequest">
  <wsdl:part element="tns:UpdatePidRequest" name="UpdatePidRequest" />
    </wsdl:message>
- <wsdl:message name="GetPidByAttributeResponse">
  <wsdl:part element="tns:GetPidByAttributeResponse"
     name="GetPidByAttributeResponse" />
    </wsdl:message>
- <wsdl:message name="GetPidByAttributeRequest">
  <wsdl:part element="tns:GetPidByAttributeRequest"
     name="GetPidByAttributeRequest" />
    </wsdl:message>
- <wsdl:message name="GetPidRequest">
  <wsdl:part element="tns:GetPidRequest" name="GetPidRequest" />
    </wsdl:message>
- <wsdl:message name="CreatePidResponse">
  <wsdl:part element="tns:CreatePidResponse" name="CreatePidResponse" />
    </wsdl:message>
- <wsdl:message name="GetHopePidResponse">
  <wsdl:part element="tns:GetHopePidResponse" name="GetHopePidResponse" />
    </wsdl:message>
- <wsdl:message name="UpdatePidResponse">
  <wsdl:part element="tns:UpdatePidResponse" name="UpdatePidResponse" />
    </wsdl:message>
- <wsdl:message name="GetPidResponse">
```

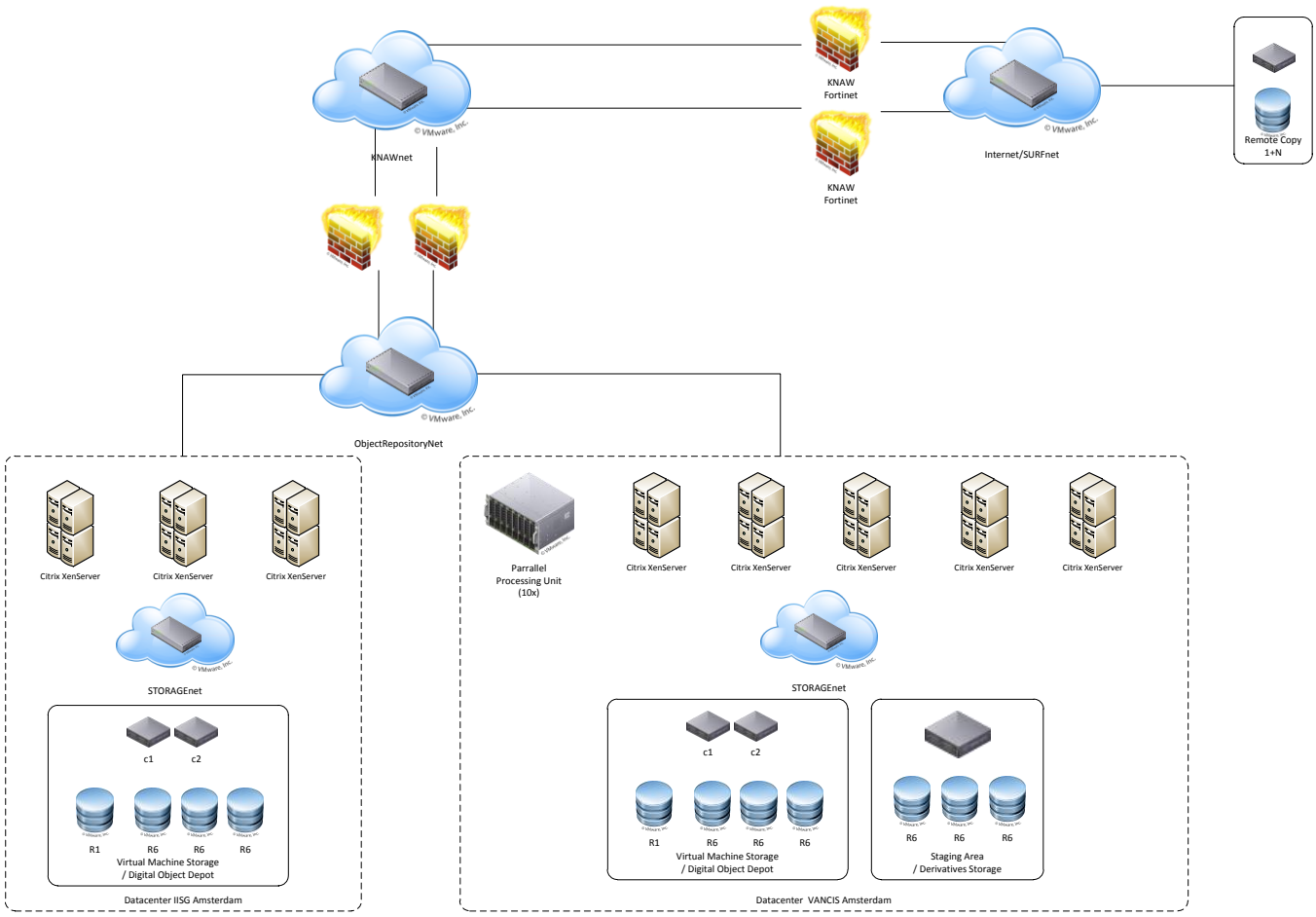
```
  <wsdl:part element="tns:GetPidResponse" name="GetPidResponse" />
    </wsdl:message>
- <wsdl:portType name="HandleResource">
- <wsdl:operation name="CreatePid">
  <wsdl:input message="tns:CreatePidRequest" name="CreatePidRequest" />
  <wsdl:output message="tns:CreatePidResponse" name="CreatePidResponse" />
    </wsdl:operation>
- <wsdl:operation name="GetHopePid">
  <wsdl:input message="tns:GetHopePidRequest" name="GetHopePidRequest" />
  <wsdl:output message="tns:GetHopePidResponse" name="GetHopePidResponse" />
    </wsdl:operation>
- <wsdl:operation name="UpdatePid">
  <wsdl:input message="tns:UpdatePidRequest" name="UpdatePidRequest" />
  <wsdl:output message="tns:UpdatePidResponse" name="UpdatePidResponse" />
    </wsdl:operation>
- <wsdl:operation name="GetPidByAttribute">
  <wsdl:input message="tns:GetPidByAttributeRequest"
     name="GetPidByAttributeRequest" />
  <wsdl:output message="tns:GetPidByAttributeResponse"
     name="GetPidByAttributeResponse" />
    </wsdl:operation>
- <wsdl:operation name="GetPid">
  <wsdl:input message="tns:GetPidRequest" name="GetPidRequest" />
  <wsdl:output message="tns:GetPidResponse" name="GetPidResponse" />
    </wsdl:operation>
    </wsdl:portType>
- <wsdl:binding name="HandleResourceSoap11" type="tns:HandleResource">
  <soap:binding style="document" transport="http://schemas.xmlsoap.org/soap/http"
     />
- <wsdl:operation name="CreatePid">
  <soap:operation soapAction="" />
- <wsdl:input name="CreatePidRequest">
  <soap:body use="literal" />
    </wsdl:input>
- <wsdl:output name="CreatePidResponse">
  <soap:body use="literal" />
    </wsdl:output>
    </wsdl:operation>
- <wsdl:operation name="GetHopePid">
  <soap:operation soapAction="" />
- <wsdl:input name="GetHopePidRequest">
  <soap:body use="literal" />
    </wsdl:input>
- <wsdl:output name="GetHopePidResponse">
  <soap:body use="literal" />
    </wsdl:output>
    </wsdl:operation>
- <wsdl:operation name="UpdatePid">
  <soap:operation soapAction="" />
```

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>
- <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:binding>
- <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>
```

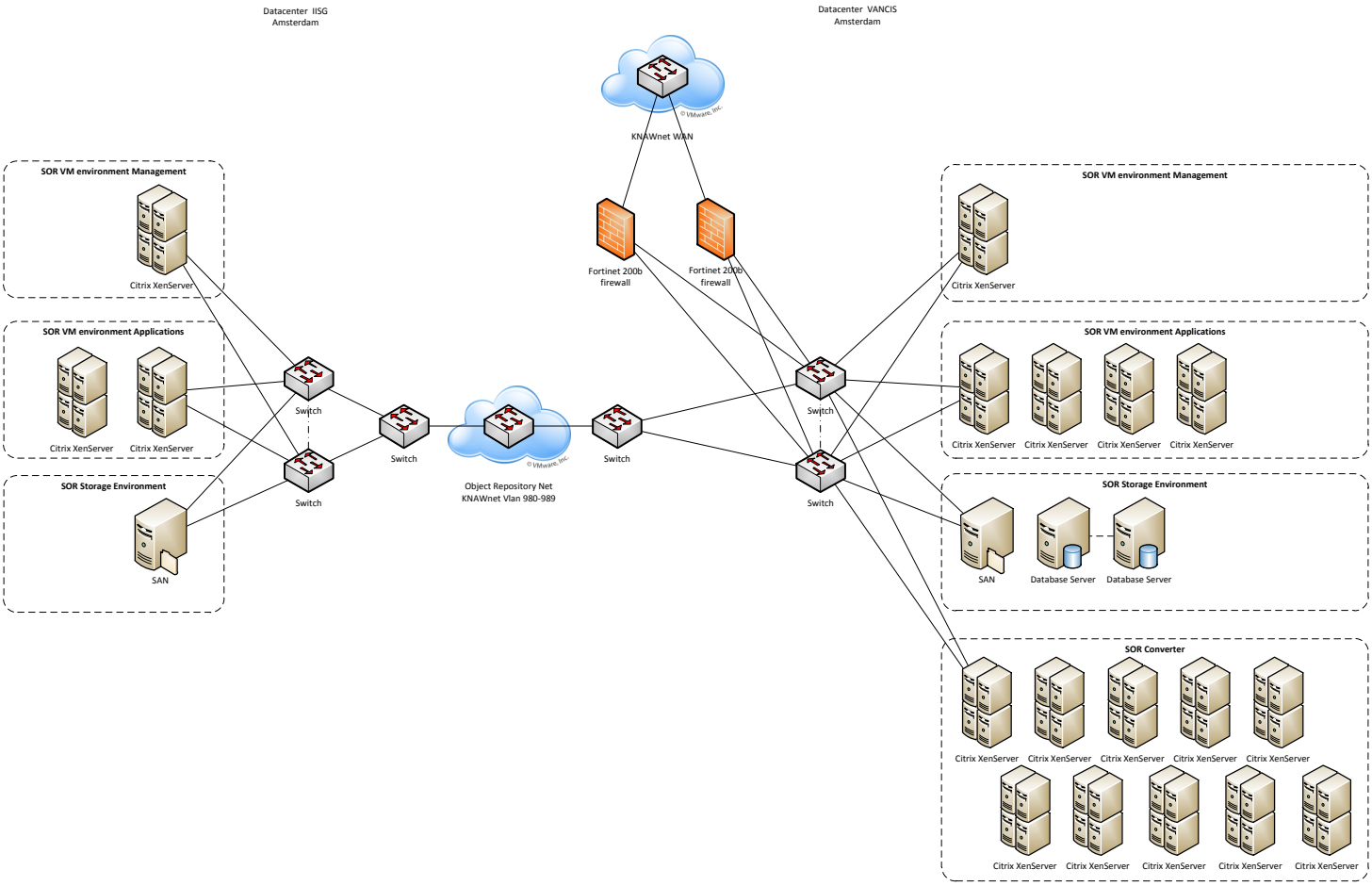
Appendix B – Low Level Design

Low level design overview



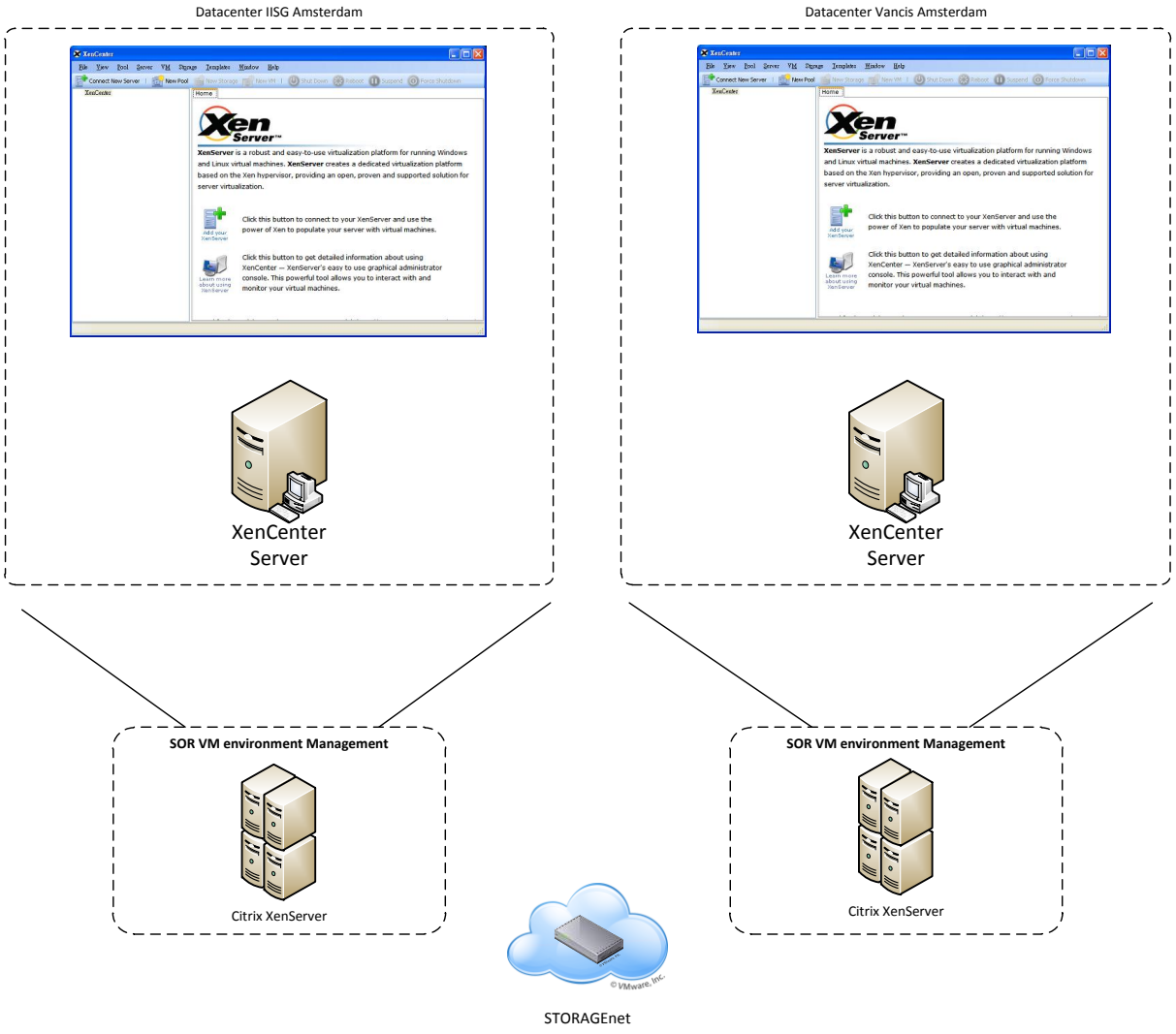
HOPE is co-funded by the European Union through the ICT Policy Support Programme.

Low level design specified

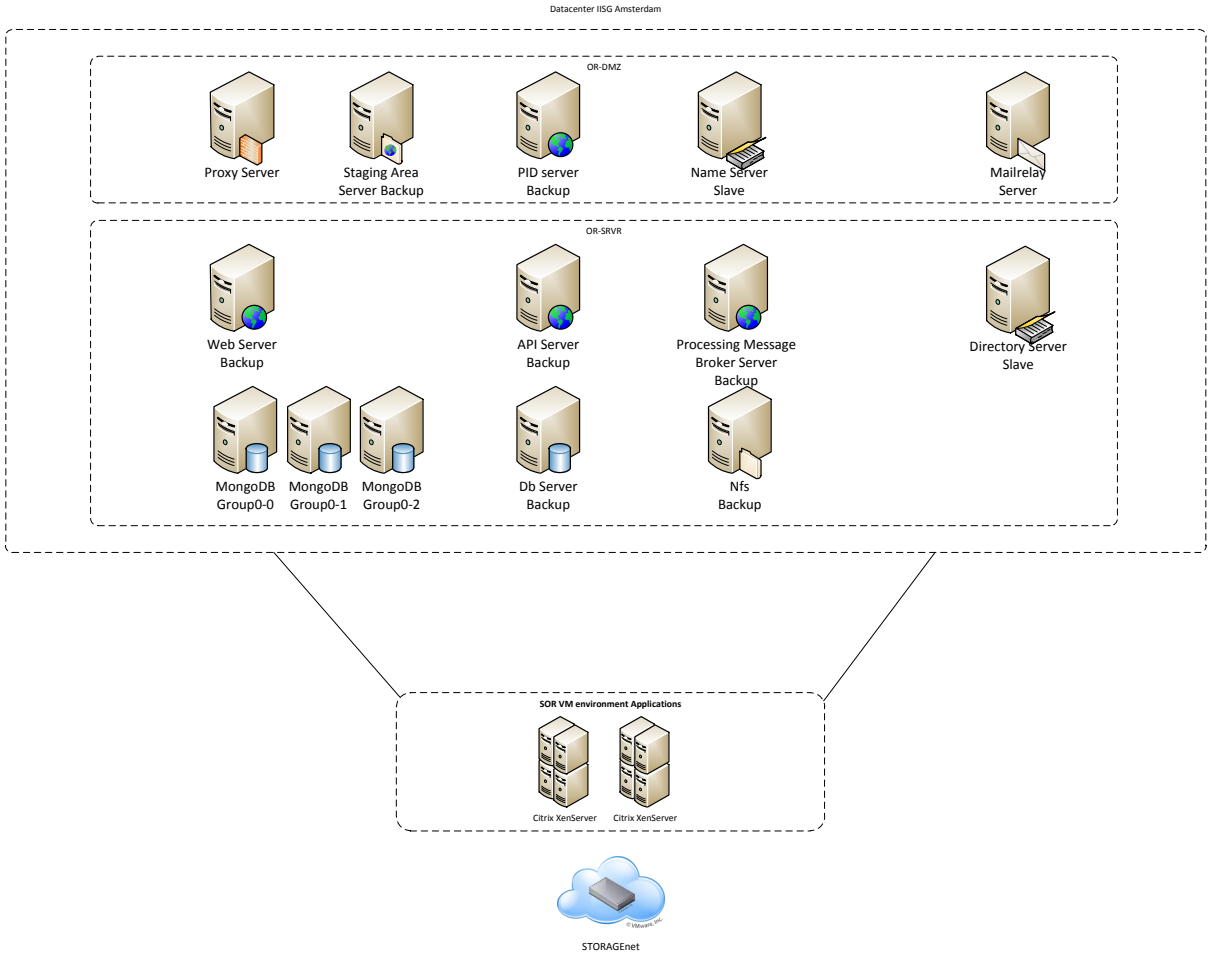


HOPE is co-funded by the European Union through the ICT Policy Support Programme.

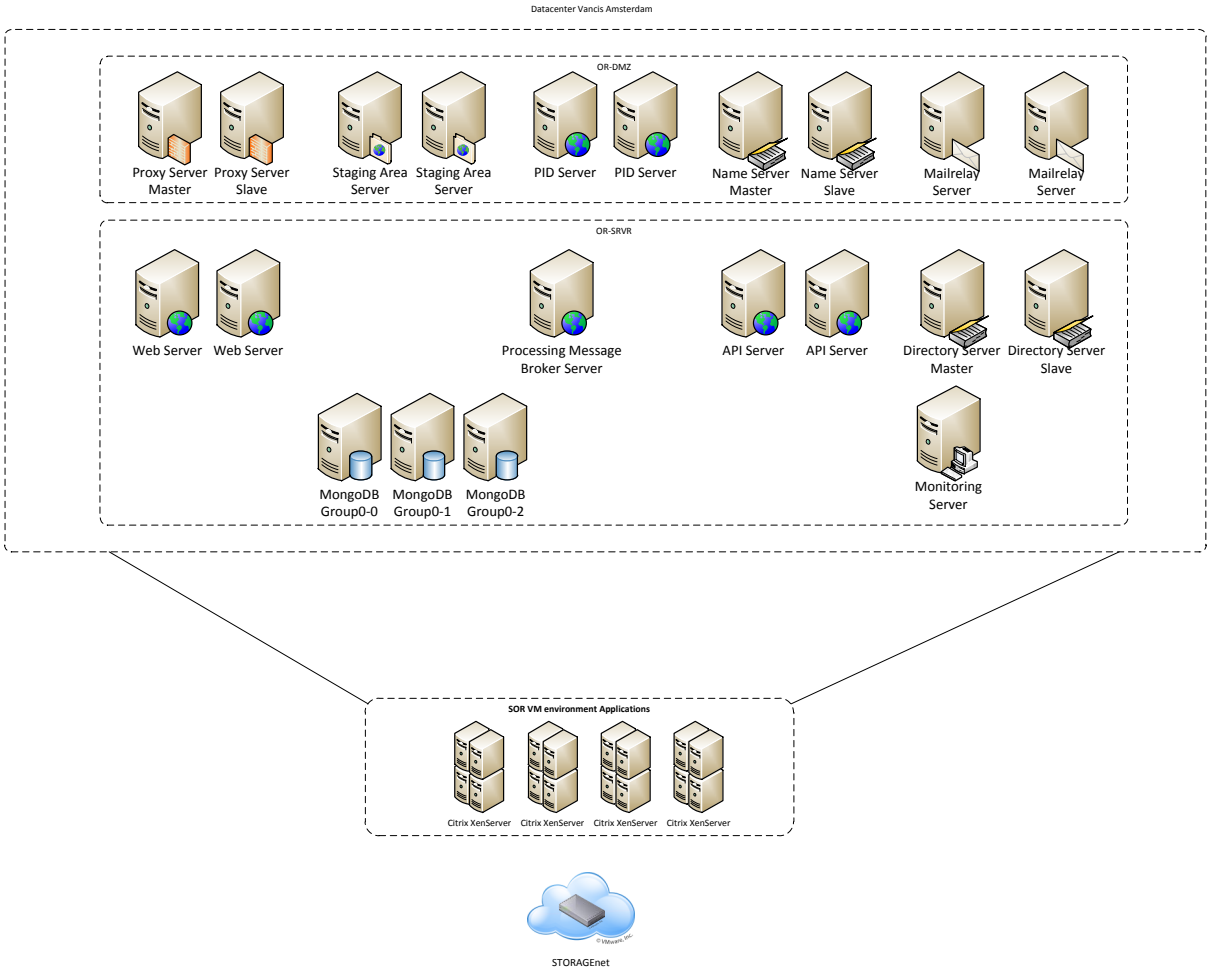
Virtual Environment Management



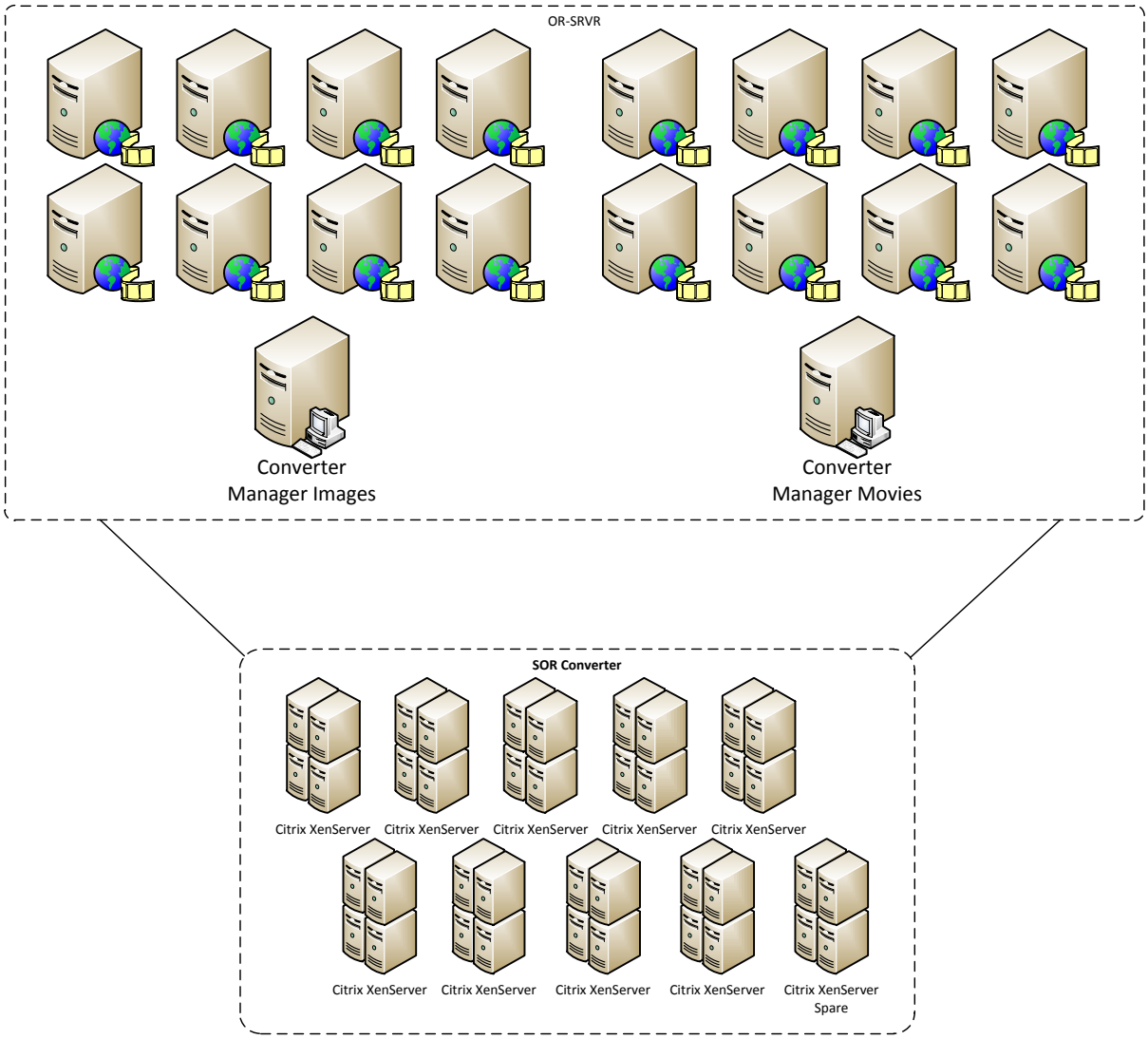
VM environment applications datacenter IISG



VM environment application datacenter Vancis

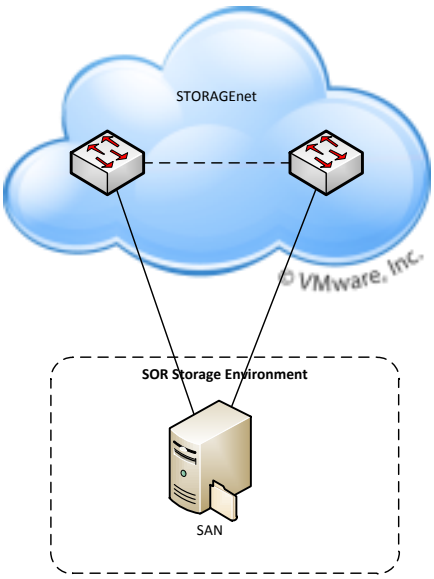


VM environment converter

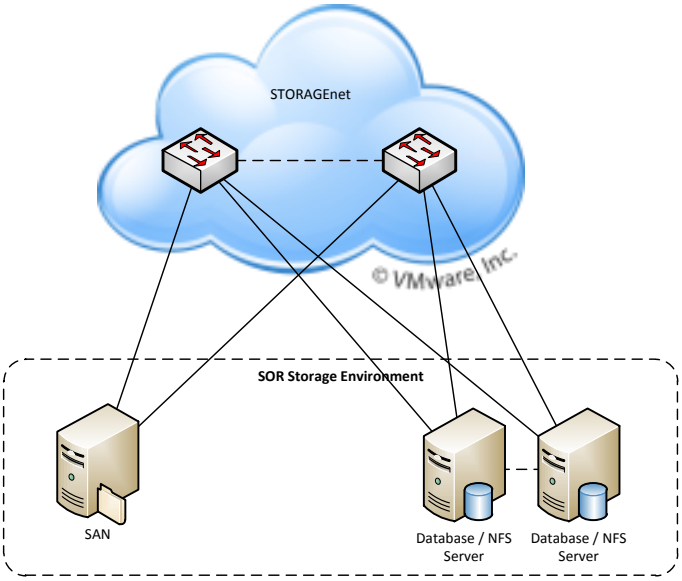


VM storage environment

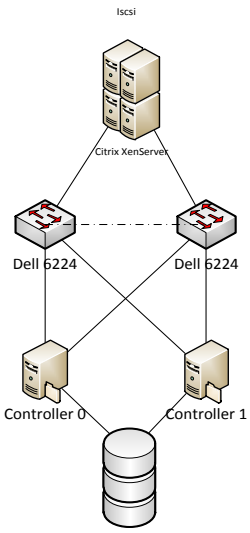
Datacenter IISG Amsterdam

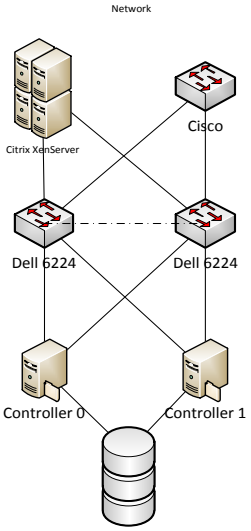


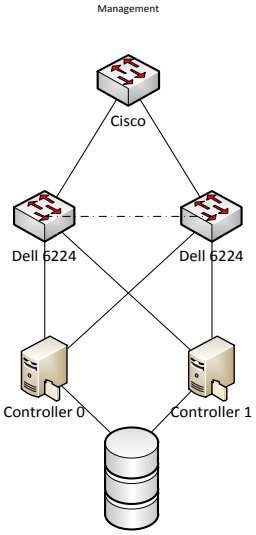
Datacenter Vancis Amsterdam

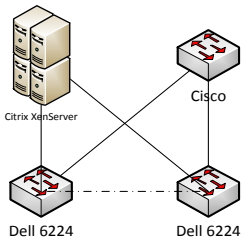


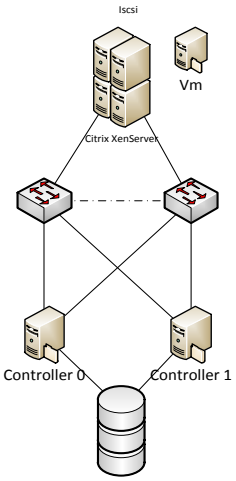
Network overview











6 NICs
2 management
2 Iscsi
2 network

or

6 NICs
1 management
1 vMotion
2 iscsi
2 network (includes backup Vmkernel for management)

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

Access Rights	The information that identifies the legal access restrictions pertaining to the *HOPE Social History Resource*, relating to legal frameworks such as the Copyright Laws, Privacy Law, etc. and licensing agreements between CPs and rights owners.	Source: HOPE See also: *HOPE Access Conditions Matrix*
Administration Interface	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.	Source: HOPE See also: Administration panel
Administration Panel	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.	Source: HOPE See also: Administration interface
API	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	Source: Wikipedia
Checksum	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	Source: File Research Center

Content Provider	A HOPE partner with social history collections which provides metadata and *Digital Objects* to the *HOPE System*.	Source: Glossary v2.0
Datacenter	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	Source: Wikipedia
DBMS	DataBase Management System; A set of computer programs that controls the creation, maintenance, and the use of a database.	Source: Wikipedia
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

Derivative Level 2	Medium to low-resolution Derivative for online consultation (view/listen) purposes.	Source: Glossary v2.0 See also: Derivative 2
Derivative Level 3	Preview-quality Derivative (lowest resolution) for display purposes in search results.	Source: Glossary v2.0 See also: Derivative 3
Derivative Storage	Storage component for derivatives in the HOPE SOR	Source: HOPE See also: MongoDB
Dev Lan	Development Local Area Network	Source HOPE See also: Local Area Network
Digital Object Depot	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	Source: Glossary v2.0 See also: Degital Object Repository
Discovery Service	A web portal, which enables the discovery, identification, and selection of materials through searching and browsing functions.	Source: Gloassary v2.0
DMZ	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.	Source: Wikipedia

Drupal	Free and open source content management system (CMS) and Content Management framework (CMF) written in PHP and distributed under the GNU General Public License	Source: Drupal.org
Event driven architecture	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.	Source: Wikipedia
Firewall	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	Source: Webopedia.com
Handle	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	Source: handle.net

	implementation of the protocols. 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.	Source: Glossary v2.0
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*.	Source: Glossary v2.0
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	Source: HOPE
HOPE Technical	Storage component of the SOR	Source: HOPE

Metadata Storage	where will be stored after ingest	See also: Technical Metadata (Glossary v2.0)
Identification, Authentication, Authorization (IAA)	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	Source: HOPE
Ingest	Process whereby, based on the SOR processing instructions, digital objects will be moved from the Staging Area to the SOR Digital Object Repository	Source: HOPE
ISCSI	Internet Small Computer System Interface; A protocol that serializes SCSI commands and converts them to TCP/IP	Source: Pcmag.com
Local Area Network (LAN)	A communications network that serves users within a confined geographical area. The "clients" are the user's workstations typically running Windows, although Mac and Linux clients are also used. The "servers" hold programs and data that are shared by the clients. Servers come in a wide range of sizes from Intel-based servers to mainframes.	Source: Pcmag.com
MongoDB	A scalable, high-performance, open source, document-oriented database.	Source: Mongodb.org

Object Oriented Architecture	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.	Source: Businessdictonary.com
Persistent Identifier (PID)	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.	Source: Glossary v2.0
RDBMS	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.	Source: Wikipedia See also: DBMS
Router	A device that forwards data packets across computer networks. Routers perform the data "traffic directing" 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	Source: Wikipedia
SAN	Storage Area Network; A network of storage disks. In large enterprises, a SAN connects multiple servers to a	Source: Pcmag.com

	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.	
Server Lan	A local area network (LAN) server is a program (and by implication usually the computer it runs in) that "serves" 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

Virtualization	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.	Source: techterms.com
VLAN	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	Source: HOPE
WAN	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.	Source: HOPE
Webservice	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	Source: W3C See also: WSDL

	standards.	
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