



VPH-Share

Period 3 Review: Online System Access

Background

VPH-Share is a 20-Partner FP7 Integrated Project that is constructing an online, cloud-based environment in which users can share software tools, applications and data, and can develop and execute *in silico* clinical workflows. The VPH-Share infrastructure has been operational for over 12 months, with new features and content being added continuously. Four Flagship Workflows are included as demonstrators and, in recent months a significant number of additional Use Cases have been added to the project, several of which are already active. Notably, a related FP7 VPH project 'MySpine' has had its entire pre-processing workflow installed in a VPH-Share virtual machine, enabling the MySpine consortium to process 200 clinical cases without the need for complex local software installation.

Reviewers: A Guided Introduction

At the project's second periodic review in April 2013 the Reviewers proposed that, to obtain an enriched appreciation of the system, they might have access to the infrastructure during the two week period prior to the 3rd Periodic Review. This has been arranged, and VPH-Share accounts have been enabled for each reviewer. VPH-Share, although carefully designed, is none-the-less a complex infrastructure and, as with all such systems, the full benefits emerge after a degree of familiarisation. To help with this learning process we have prepared a set of illustrated guidance slide-sets that show users how to begin using the system and carry out three common activities.

For the Busy Reviewer: In all cases you'll need to see through the Login process. The three activities are then described in separate tutorials, and each requires around 40 minutes to complete. If time doesn't permit this degree of engagement, it may be that an agreement between Reviewers to split the three tasks amongst you may be a possible solution. If this is palatable, the data-centred activities are probably the least challenging (perhaps 4/10), with the application and workflow development activities being the most challenging (perhaps 7/10)

If you're particularly time-challenged, but would still like to try the system, the Data activity includes as its final step a search exercise to identify semantically-tagged data; this will give a flavour of the system without requiring a large investment of time.

Guided Activity #0: Login

To avoid the irritation that can arise from initial engagement with a new system, this first **Login** guide takes the user carefully through the Login process. In full, the process would be...

- a. Locate the VPH-Share Portal <https://portal.vph-share.eu/> (also easily found through Google)
- b. Register on the VPH-Share site (already done for Reviewers)
- c. Obtain an account on the central VPH authentication system Biomed Town (already done)
- d. When notified of your account activation, revisit the portal and Login
- e. Subscribe to any relevant group (Reviewers are already in the 'Reviewers' group)
- f. (Note the online help and status information)

Only step 4 is strictly required for Reviewer access (though steps 5 and 6 are worth a look).



Guided Activity #1: Data Handling

VPH-Share considers data in two distinctly different categories:

- Unstructured data – data in individual files, lacking a known or exposed schema describing its organisation – is stored in a conventional ‘Filestore’, and is tagged with a small number of attributes that apply at the level of the entire data file only.
- Structured data (typically tabulated information) provides the additional opportunity for description and manipulation at a more detailed level of granularity (categories: ‘fields’ or ‘columns’, and records: ‘entries’ or ‘rows’) and is supported by a series of VPH-Share facilities. A special tool, the VPH-Share Data Publication Suite, is available for use both as a cloud-based Application, and as a software component that can be installed locally on a user’s Windows PC (typically behind a hospital firewall), and provides facilities to assist both the data provider and other potential data users by facilitating enhancement of the data prior to upload to the VPH-Share infrastructure.

The handling of unstructured data in the VPH-Share Filestore is very straightforward, so this **Data** guide instead provides an exploration of the handling facilities for structured data:

- a. The user installs the Data Publication Suite (DPS) locally and obtains an example dataset containing several related files
- b. Using the DPS, one of the dataset files is opened as a data ‘source’, and a previously-established location within the VPH-Share system is identified as a ‘destination’.
- c. Particular fields within the source can now be given special attention:
 - One field can be set as containing an identifier for each entry
 - Fields that contain links to other, related, data files (such as images) can be set up to have the relationship recognised by the VPH-Share system
 - Fields can be tagged semantically (with reference to known ontologies) such that the content is systematically and formally identifiable and linkable using Resource Description Framework (RDF) tags. This also generates the ‘metadata’ for the data resource – the formalised description of the content that can assist others searching for possible data sources
 - Fields can be marked such that their contents are altered prior to upload, or omitted
 - In particular, sensitive fields can be marked such that the content is pseudonymised using an industry-standard obfuscation method
- d. When all conversion parameters have been established, the settings can be saved and the data uploaded
- e. The resulting dataset can be queried using the standard formalised RDF query language ‘SPARQL’, which operates on the RDF tags to provide a semantics-based interrogation facility.
- f. The existence of the published data can be made known, and the data itself can be shared with other trusted users. Once the data’s existence is exposed, other VPH-Share users can interrogate the metadata (but not the data itself) to determine its suitability. [This item represents basic VPH-Share usage, and might be attempted by the time-pressed Reviewer].



Guided Activity #2: Application Development

VPH-Share provides for the development and sharing of computational processes that are useful as components of larger workflows (and perhaps also as stand-alone entities). This **Application** guide covers the creation of a VPH-Share Application built from a pre-existing service.

- a. Become familiar with the process of using VPH-Share Applications
 - Find an existing Application already available within VPH-Share, note its characteristics
 - Start an instance of the Application
 - Identify important information required to make use of this now-running Application
- b. Create a new Application from an existing program (which is provided to you); first, install the program and try it out on your local PC.
- c. Create the VPH-Share Application...
 - Obtain the necessary confidential 'Secure Shell' SSH credentials and uploading them
 - Choose the basic 'wrapper' that will be used to encapsulate the Application
 - Define the details of the wrapper, including its 'endpoint' (the location where the service will become available)
- d. Make the Application available to other users.

Guided Activity #3: Workflow Development

The ultimate goal of many VPH-Share users is to employ the Infrastructure to develop complete workflows capable of accepting patient data as inputs and presenting clinically-relevant diagnostic or treatment information as outputs. These workflows are typically assembled from pre-existing Applications that are chained together to interact and provide clinical solutions. This **Workflow** guide covers the creation of such a workflow:

- a. We start with an incomplete workflow that will be finished by the addition of missing facilities
- b. The incomplete workflow is opened in the online workflow editor 'Taverna Online', which gives a diagrammatic view of the existing components and their interconnections
- c. A new service is added, its parameters are entered, and it is connected into the existing workflow
- d. The resulting workflow is saved with a new name
- e. We execute the new workflow, providing details of the required source data
- f. We inspect various details of the running workflow 'instance'
- g. Finally we share the completed and tested workflow with other users