

VPH-Share Research Platform

- 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.
- 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
- Operational for over 12 months, with new features and content being added continuously.

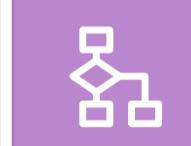


© VPH-Share 2014

Period 3 Review, 23-May-14

Email support@vph-share.eu

2

			
Unified Identity	Security	Workflows	Semantics
			Structured Data
High Performance	Applications		
		Cloud Computing	Unstructured data
<h2>Platform Building Blocks</h2>			
 © VPH-Share 2014	Period 3 Review, 23-May-14	Email support@vph-share.eu	3

VPH-Share - Research Partnerships



© VPH-Share 2014

Period 3 Review, 23-May-14

Email support@vph-share.eu

4

Slide Objectives:

- Describe our current mix of partners and collaborators

Speaking Points:

Reviewers: 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.



© VPH-Share 2014

Period 3 Review, 23-May-14

Email support@vph-share.eu

5

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.



© VPH-Share 2014

Period 3 Review, 23-May-14

Email support@vph-share.eu

6



Summary

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.



© VPH-Share 2014

Period 3 Review, 23-May-14

Email support@vph-share.eu

8

Summary

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



© VPH-Share 2014

Period 3 Review, 23-May-14

Email support@vph-share.eu

9

Summary

- When all conversion parameters have been established, the settings can be saved and the data uploaded
- 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.
- 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].



© VPH-Share 2014

Period 3 Review, 23-May-14

Email support@vph-share.eu

10

Data Consumer & Publisher

01 Data Publication Suite

02 Data Import & Transformation

03 Semantic Annotation

04 Data De-identification

05 Data Publication & Query

06 Sharing & Permissions

07 Data Search & Discovery



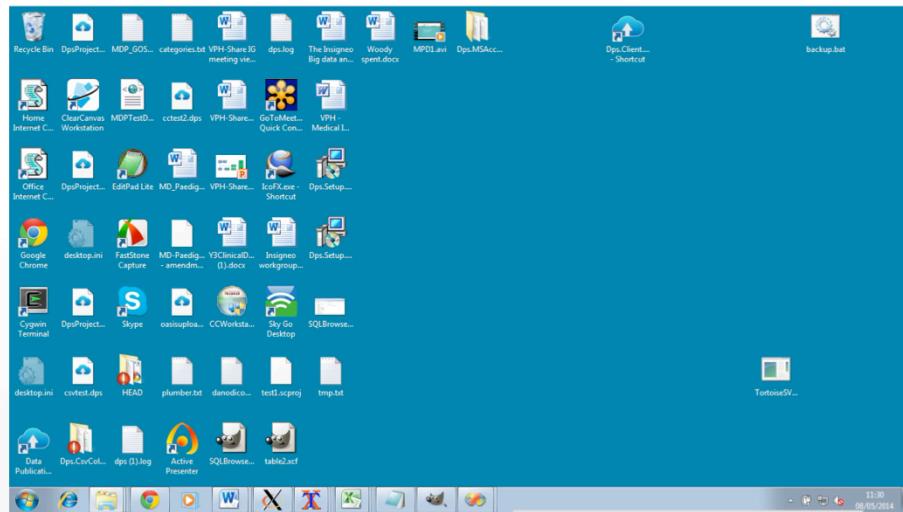
© VPH-Share 2014

Period 3 Review, 23-May-14

Email support@vph-share.eu

11

Video - Data Publishing



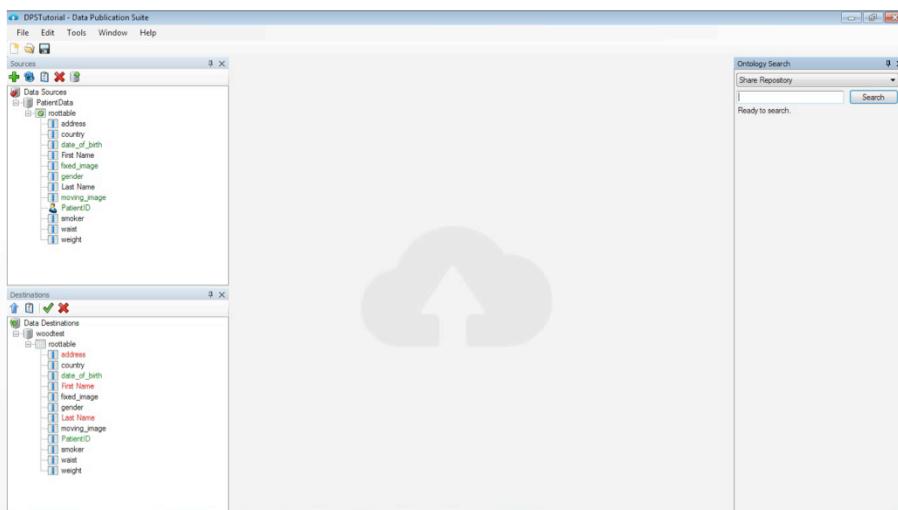
© VPH-Share 2014

Period 3 Review, 23-May-14

Email support@vph-share.eu

12

Video - Data Permissions & Metadata



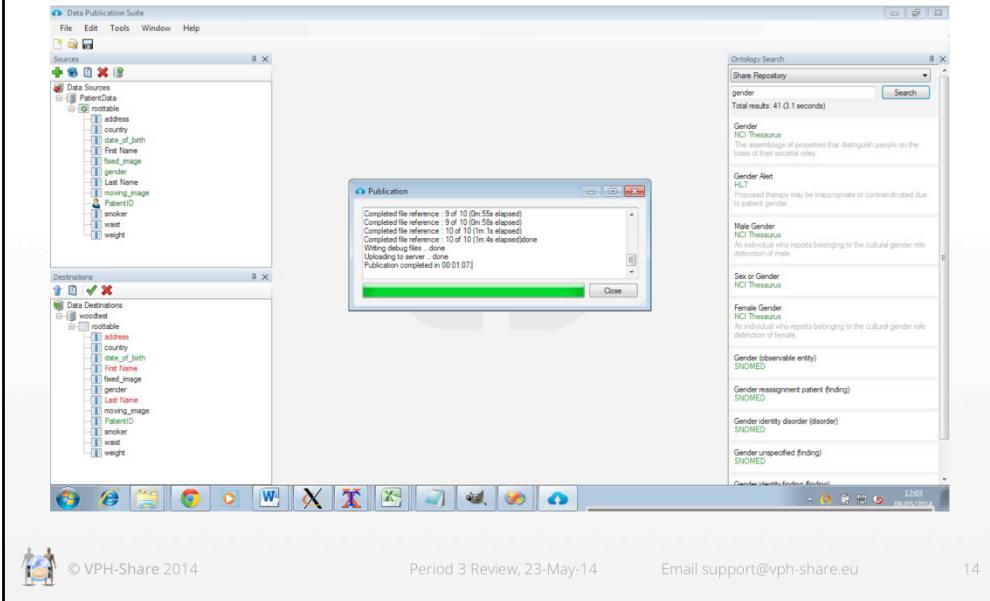
© VPH-Share 2014

Period 3 Review, 23-May-14

Email support@vph-share.eu

13

Video - Data Post Publishing



01 Data Publication Suite

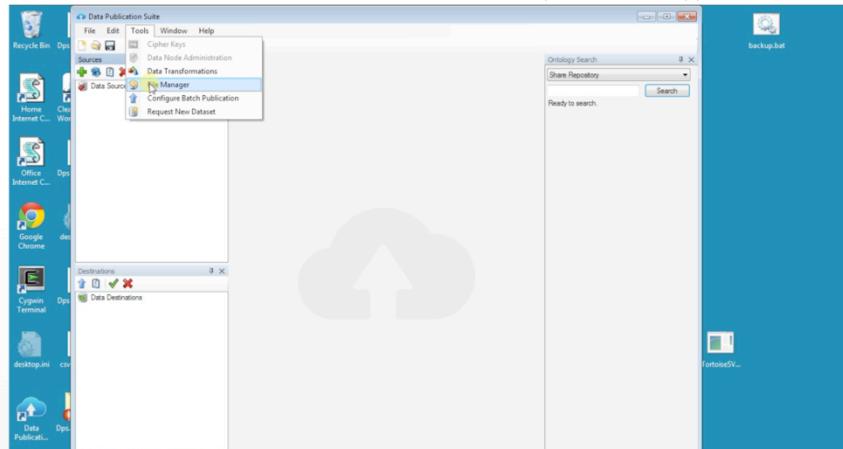
- a. Click 'Data' and download 'Data Publications Suite' for your architecture (64-bit or 32-bit)
- b. Double-click and install 'Data Publication Suite'
- c. Download "Test Dataset" & extract to C:/ [IMPORTANT]

The screenshot shows the VPH-SHARE Data Publication Suite interface. At the top, there's a navigation bar with links for Help, About, and a user profile (Hi, Susheel). Below the navigation is a banner titled "Biomedical Data at your Fingertips" featuring a grid of brain scan images. The main content area has several sections: "Browse Data" (with a search icon), "Search Data" (with a search icon), "Upload Data" (with a cloud icon), and "Documents" (listing User's manual, Data Sharing Protocol, and Terms of use) and "Applications" (listing DPS installer (64-bit), DPS installer (32-bit), DPS Test Dataset, and FileStore). A sidebar on the left lists "DATA", "APPLICATIONS", "WORKFLOWS", and "SEARCH". At the bottom, there's a footer with copyright information (© VPH-Share 2014), a review date (Period 3 Review, 23-May-14), an email address (support@vph-share.eu), and a page number (15).

User Manual: <http://www.vph-share.eu/content/structured-data-data-publication-suite>

01 Data Publication Suite

- a. Click 'Data' and download 'Data Publications Suite' for your architecture (64-bit or 32-bit)
- b. Double-click and install 'Data Publication Suite'
- c. Download "Test Dataset" & extract to C:/ [IMPORTANT]
- d. Double-click desktop icon to start application, maximise



VPH-Share provides a fully integrated data publication application that provides data import, transform, annotation, de-identification, publication and query components.



© VPH-Share 2014

Period 3 Review, 23-May-14

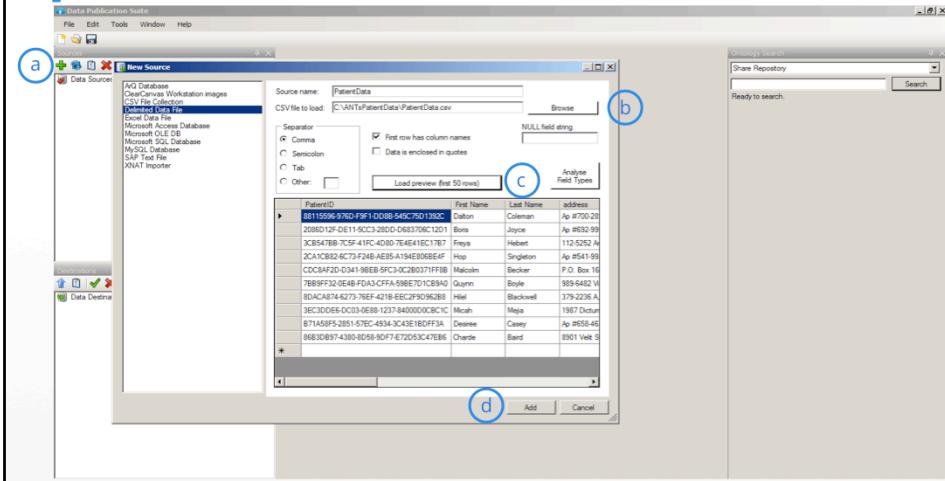
Email support@vph-share.eu

16

User Manual: <http://www.vph-share.eu/content/structured-data-data-publication-suite>

02 Data Import & Transform

- a. Click 'New Data Source' and select 'Delimited Data File'
- b. Click 'Browse' and select 'PatientData.csv' in C:\ANTs...
- c. Optionally Click 'Load Preview'
- d. Click 'Add' to add new data source



The Data publication suite supports multiple structured data formats that may be transformed, annotated and published to the VPH-Share network



© VPH-Share 2014

Period 3 Review, 23-May-14

Email support@vph-share.eu

17

User Manual: <http://www.vph-share.eu/content/connect-data-source>

02 Data Import & Transform

- e. Open 'PatientData' and 'roottable' in data source tree
- f. Right-click 'roottable' and select 'View Data Types'
- g. Right-click 'roottable' and select 'View Table Contents'

The screenshot shows the Data Publication Suite application window. On the left, the 'Sources' pane displays a tree structure with 'PatientData' expanded, showing 'roottable' selected. A circled 'e' is next to 'roottable'. Below it are 'Set as key', 'Rename', 'Relationships', 'Properties', 'View Data Types', and 'View Table Contents'. Another circled 'f' is next to 'View Data Types'. The central part of the screen shows the 'FieldAnalysis' dialog with a table of data types for 'date_of_birth' (date) and 'smoker' (int). To the right is the 'Results Viewer roottable' dialog, which contains a table of patient data:

PatientID	First Name	Last Name	address	country	weight	rest	date_of_birth	smoker
83115596-97D0-	Delon	Coleman	Ap #700-2897 D.	Virgin Islands, Br.	51.9559226	38.65964957	12/01/1980	0
2086012F-0E11-	Boris	Joyce	Av #930-9994 C.	Samoa	57.92852995	33.78995416	22/01/1982	0
3C847B8B-7C5F-	Fraya	Hebert	112-632-0vee 5.	Slovenia	62.23324131	32.1576081	17/09/1985	1
2CA1C820-8C73-	Hop	Singleton	Ap #541-9922 P.	Benin	50.53191368	34.5385665	09/09/1977	0
CDC4A920-D34-	Malcolm	Becker	P.O. Box 168-54.	Guinea-Bissau	45.21699134	36.0445105	23/06/1981	1
7B89PF32-0E4B-	Quynn	Boyle	935-6482 Vivera,	Turkmenistan	50.99911053	39.0572724	24/04/1989	1
8D4CA37A-6273-	Hiel	Blackwell	375-2236 A, St.	Jordan	42.97457525	39.14695053	18/02/1979	1
3EC30DE6-0C0-	Micah	Mejia	1987 Dicum Rd.	Taiwan	45.5598206	35.2576685	04/10/1983	0
871A59F5-2851-	Desree	Casey	Ap #858-4531 Pr.	United Kingdom, L...	60.84924424	35.40281641	08/02/1981	1
86830897-4380-	Chardie	Bard	8901 Velt Street	Nepal	57.74204958	37.24858955	06/01/1975	1

The Data publication suite provides simple table analysis components to preview the data imported and sanity check the tables data-types.



© VPH-Share 2014

Period 3 Review, 23-May-14

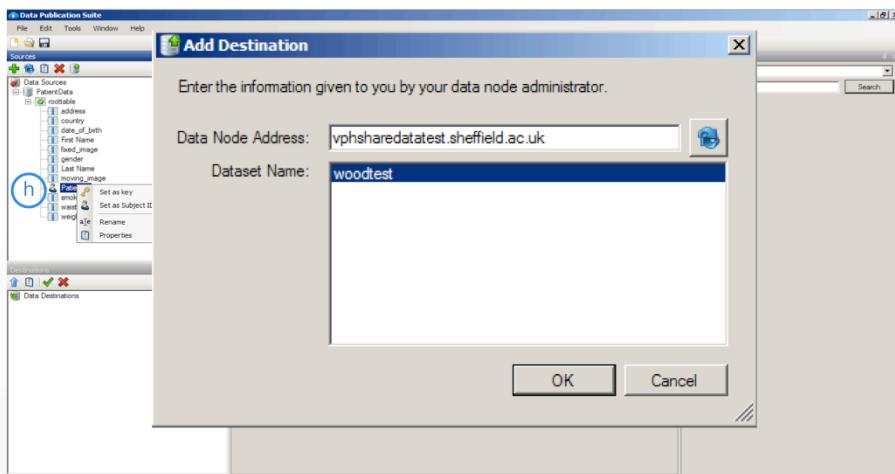
Email support@vph-share.eu

18

User Manual: <http://www.vph-share.eu/content/connect-data-source>

02 Data Import & Transform

- h. Right-click 'Patient ID' column and 'Set Subject ID'
- i. Right-click 'fixed_image' and 'moving_image', select Properties > Check 'File reference'



The Data publication suite provides multiple data transformations that can be applied via plugins. Users are free to add their own data transformation plugins.



© VPH-Share 2014

Period 3 Review, 23-May-14

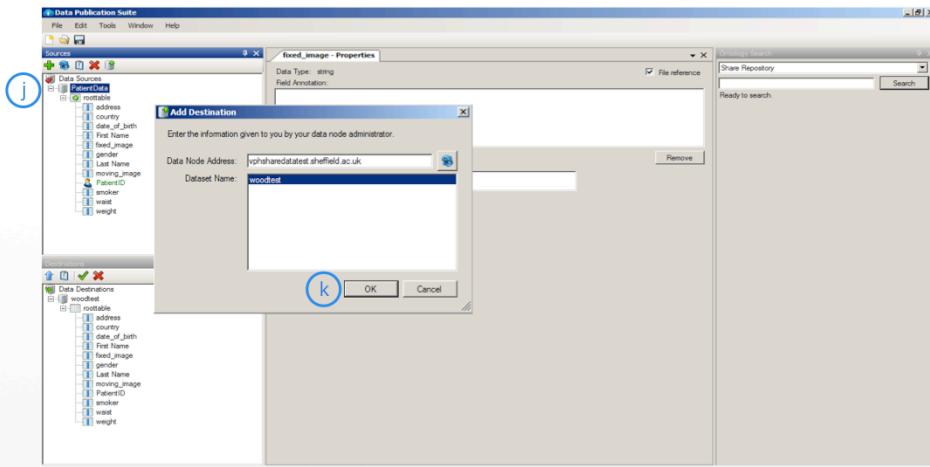
Email support@vph-share.eu

19

User Manual: <http://www.vph-share.eu/content/working-tables-and-fields>

02 Data Import & Transform

- h. Right-click 'Patient ID' column and 'Set Subject ID'
- i. Right-click 'fixed_image' and 'moving_image', select Properties > Check 'File reference'
- j. Right-Click 'PatientData' & select 'Add New Destination'
- k. Click Refresh, Enter credentials, choose container



The Data publication suite provides multiple data transformations that can be applied via plugins. Users are free to add their own data transformation plugins.



© VPH-Share 2014

Period 3 Review, 23-May-14

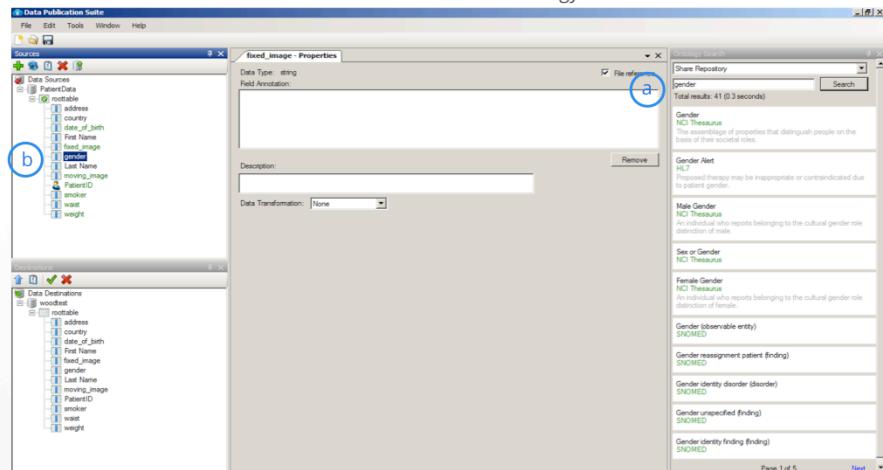
Email support@vph-share.eu

20

User Manual: <http://www.vph-share.eu/content/working-tables-and-fields>

03 Semantic Annotation

- Type 'gender' in Ontology Search box
- Drag and drop **ontology result** from results window onto the 'gender' column, which will turn **green**
- Continue annotating other columns like 'date_of_birth' with 'dob' in ontology search



VPH-Share provides a curated set of biomedical ontologies that can be used to annotate dataset columns to enable users to search using semantic concepts.



© VPH-Share 2014

Period 3 Review, 23-May-14

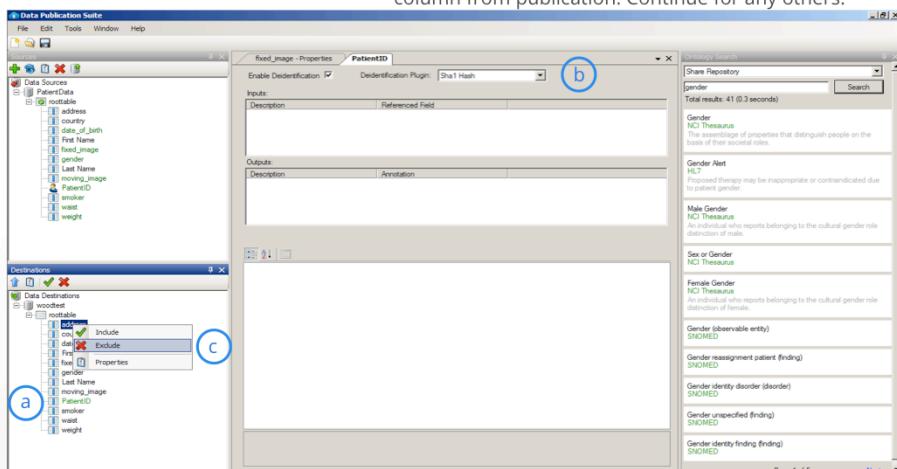
Email support@vph-share.eu

21

User Manual: <http://www.vph-share.eu/content/annotation>

04 Data De-identification

- Right-click **PatientID destination** column > **Properties**
- Check **De-identification** > Choose **SHA1 hash**. Other plugins also transform dates, etc.
- Right-click '**address**' destination column to **Exclude** column from publication. Continue for any others.



VPH-Share also provides tools to de-identify datasets and a filtering specific columns out of the publication process. One-way encryption is a standard plugin provided.



© VPH-Share 2014

Period 3 Review, 23-May-14

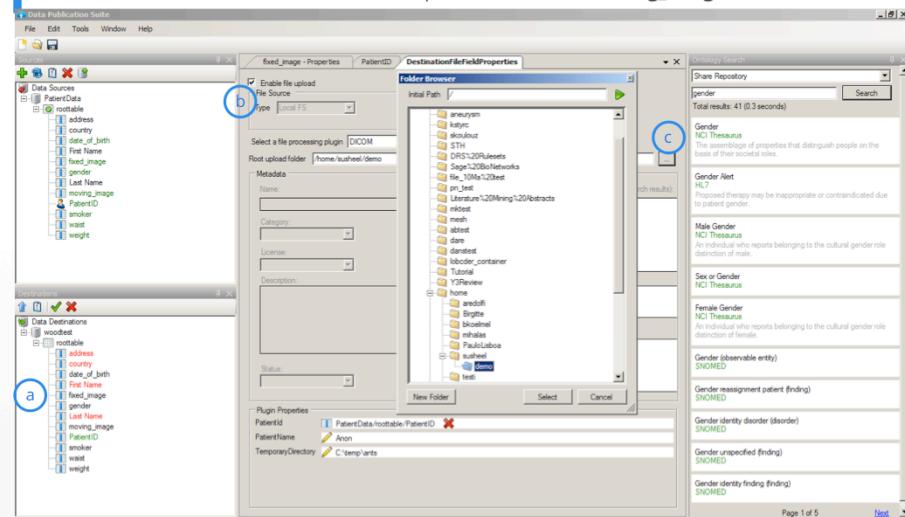
Email support@vph-share.eu

22

User Manual: <http://www.vph-share.eu/content/de-identifying-dataset>

05 Data Publication & Query

- Select **fixed_image destination** column > **Properties**
- Choose 'Enable File Upload' and 'DICOM' file processing
- Choose '**Root upload folder**' > **/home/NAME/demo**
- Repeat on the other **moving_image** column



Data Publication Suite also provides a query-builder components to obtain data from the recently published annotated dataset.

© VPH-Share 2014

Period 3 Review, 23-May-14

Email support@vph-share.eu

23

User Manual: <http://www.vph-share.eu/content/data-publication>

05 Data Publication & Query

- e. Drag & Drop '**PatientID**' source to Plugin Properties
- f. Type '**Anon**' in **PatientName**
- g. Choose a local **TemporaryDirectory** e.g. **C:/temp/ants**

VPH-Share allows users to publish their datasets to multiple destinations if needed. Data Publication Suite also provides a query-builder components to obtain data from the recently published annotated dataset.

© VPH-Share 2014 Period 3 Review, 23-May-14 Email support@vph-share.eu 24

User Manual: <http://www.vph-share.eu/content/data-publication>

05 Data Publication & Query

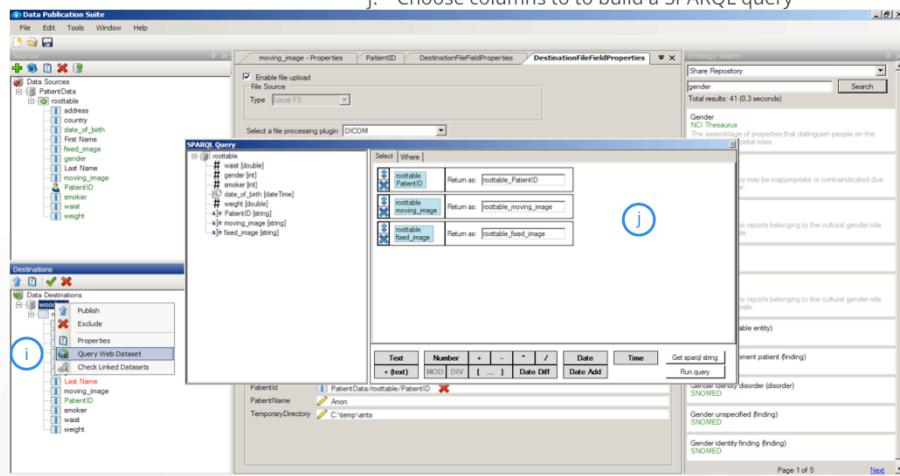
h. Right-Click Destination Container > Publish

The screenshot shows the VPH-Share Data Publication Suite interface. On the left, there's a tree view of 'Data Sources' containing 'PatientData' with various fields like 'name', 'address', 'country', etc., and 'Destinations' which includes 'Publish'. A circled 'h' is next to the 'Publish' item. In the center, a 'moving_image - Properties' dialog is open, showing 'DestinationfileFieldProperties'. A 'DestinationfileFieldProperties' dialog is also visible. A 'Share Repository' window on the right lists terms like 'gender', 'Male', 'Female', etc. A 'Publication' progress dialog shows 'Processing done' and 'Completed file reference: 1 of 10 (0m:29s elapsed)'. At the bottom, there are copyright information ('© VPH-Share 2014'), a date ('Period 3 Review, 23-May-14'), an email address ('Email support@vph-share.eu'), and a page number ('25').

User Manual: <http://www.vph-share.eu/content/data-publication>

05 Data Publication & Query

- h. Right-Click **Destination Container** > **Publish**. Publication will take a few minutes.
- i. Right-Click **Destination Container** > **Query Web Dataset**
- j. Choose columns to build a SPARQL query



VPH-Share allows users to publish their datasets to multiple destinations if needed. Data Publication Suite also provides a query-builder components to obtain data from the recently published annotated dataset.

© VPH-Share 2014

Period 3 Review, 23-May-14

Email support@vph-share.eu

26

User Manual: <http://www.vph-share.eu/content/data-publication>

06 Data Sharing & Permissions

- Right-click **Destination container** > **Properties**
- Manage Metadata. E.g. Change **License** or **Category**
- Click **Manage read-only group**
- Search and add 'Reviewers' group

VPH-Share allows fine-grained permissions on all resources. Sharing may be performed on individuals and/or groups according to the needs of the resource owner.

© VPH-Share 2014 Period 3 Review, 23-May-14 Email support@vph-share.eu 27

User Manual: <http://www.vph-share.eu/content/grant-permissions>

07 Data Search & Discovery

Navigate to the VPH-Share Portal
a. Click **Search** > Type '**Patient**' > Choose Facet '**Dataset**'
b. Choose your dataset, view metadata > Query Dataset

The screenshot shows the VPH-Share Portal interface. At the top, there's a navigation bar with links for DATA, APPLICATIONS, WORKFLOWS, and SEARCH. The SEARCH tab is active, showing a search bar with 'patient' typed in. Below the search bar, it says 'We've found 11 repository results'. There are several search results listed, each with a small icon, the dataset name, and a brief description. One result, 'woodtest', is highlighted with a blue circle labeled 'b'. To the right of the search results is a sidebar with filtering options. The 'Filter By' section has a 'Dataset' facet selected, indicated by a blue circle labeled 'a'. Other facets shown include Workflow, Application, File, Semantic web service, User, and Institution. Below this is a 'Refine' section with fields for Category, Author, Licence, and Tags, each with a corresponding input field.

VPH-Share allows users to search for data and any other resources via full-text search



© VPH-Share 2014

Period 3 Review, 23-May-14

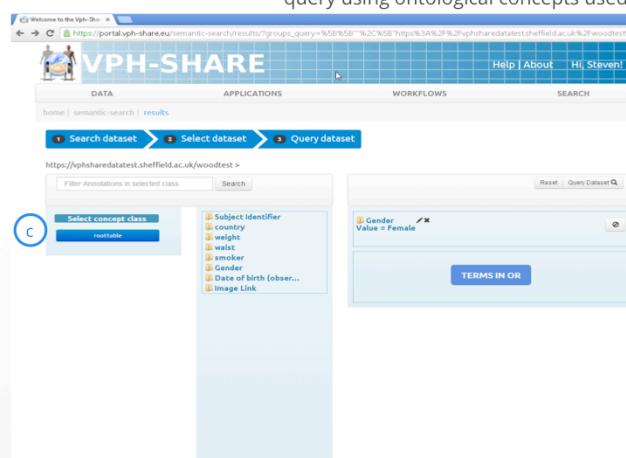
Email support@vph-share.eu

28

User Manual (General Search): <http://www.vph-share.eu/content/general-search>

07 Data Search & Discovery

- Navigate to the VPH-Share Portal
- Click **Search** > Type '**Patient**' > Choose Facet '**Dataset**'
 - Choose your dataset, view metadata > Query Dataset
 - Drag & Drop 'Semantic Concepts' to perform semantic query using ontological concepts used in annotation



VPH-Share also provides semantic query of specific datasets using ontological concepts



© VPH-Share 2014

Period 3 Review, 23-May-14

Email support@vph-share.eu

29

User Manual (Semantic Search): <http://www.vph-share.eu/content/semantic-query>

GLOSSARY

- API – Application Programming Interface
- Dataset – A collection or related data
- Delimited Data File – File format that stores data separated by a delimiter
- Data Source – Local source data from which the data publication process can be performed
- Data Destination – Destination to which data will be published
- Ontology – Represents the hierarchy of concepts within domain
- SHA1 Hash – A one-way encryption function to encrypt input
- DICOM – An image format frequently used in the medical domain
- SPARQL – A query language that is used to extract knowledge from an RDF graph



© VPH-Share 2014

Period 3 Review, 23-May-14

Email support@vph-share.eu

30

VPH-Share Consortium Partners



The
University
Of
Sheffield.



Sheffield Teaching Hospitals NHS
NHS Foundation Trust
Excellence as standard



Slide Objectives:

Speaking Points:

CONTACT US



Webpage:
vph-share.eu



Twitter:
[@vphshare](https://twitter.com/vphshare)



Email:
support@vph-share.eu



Facebook:
[VPH-Share](#)



Google+:
[VPH-Share](#)



© VPH-Share 2014

Period 3 Review, 23-May-14

Email support@vph-share.eu

32

Slide Objectives:

Speaking Points: