SPHN Semantic Interoperability Framework compliance and process for Data Transfer Requests (DTRs)

The National Data Streams (NDS) and Demonstrator Projects (DEM) are the first SPHN projects that will operate fully within the SPHN Semantic Interoperability Framework. The Project Schema creation and the process for Data Transfer Requests (DTRs) need to be done in a harmonized manner to ensure a smooth collaboration between the project teams, the central Data Coordination Center (DCC) teams, and the data-providing institutions. The following is a list of steps for the preparation of the project RDF Schema and external terminologies, de-identification rules as well as the workflow for the DTRs in SPHN projects elaborated together with representatives of the University Hospitals and SPHN projects.

Version1, 1st of September 2023

1. Preparatory steps

	Task	*Responsibility*	*Comment*
A	**Set up of Git Group** (Git group for each project)		
	Assign a maintainer to manage the project's Git group and user permission	NDS/DEM, (DCC)	
	Ensure that all necessary persons have access to the project's Git group (project members, DPs / UHs personnel, such as DP coordinators, data engineers, project coordinators, etc.)	NDS/DEM	
B	**Project RDF Schema**		
	Design semantic concepts and submit proposals to the DCC (PDF files using the Templates provided by DCC, following the Guiding principles)	NDS/DEM, (DP, RDF Support)	Can be an iterative process
	Receive feedback on concepts (Word files)	DCC	
	Implement feedback and resubmit to DCC until an agreement is reached	NDS/DEM, (RDF Support)	
	Integrate concepts into the project Dataset template (Template, User guide, Training)	NDS/DEM, (RDF Support)	
	Generate RDF (.ttl), SHACL, SPARQLs, and HTML using Schema Forge; upload documents on project Git (Git group) **Folder structure and naming** **Group** *Project Name* **Repository*: "**rdf-schema" (* Public) – **Folder:** "version-1"	NDS/DEM, (RDF Support)	A strict adherence to the folder structure is necessary so that the SPHN Connector can directly access the

	– Subfolder "*dataset*" (Excel file and PDF documents describing the concepts) – Subfolder "*schema*" (.ttl file of the Project RDF Schema) – Subfolder "*shacl*" *(optional, only needed if SHACLs are manually adapted)* – Subfolder *"sparql" (optional)* – Subfolder "*doc*" (HTMLfor website documentation of the schema) – **Folder:** "version-2" …		required files in the future
	Add tag "Release-candidate-schema-<project-name>-<year>-<version>" to the repository	NDS/DEM	
	Notify the DCC once the schema is ready for review (email to dcc@sib.swiss)	NDS/DEM	Can be an iterative process
	Review of the documents	DCC	
	Add tag "DCC-approved-schema-<project-name>-<year>-<version>"	DCC	
	Publish the HTML documentation of the schema on the biomedit.ch website	DCC	
C	**External Terminologies**		
	FAIRify and translate external terminologies used in the project to SPHN compliant RDF (User guide, Training)	NDS/DEM, (RDF Support)	
	Upload in RDF files (.ttl) on project Git (Git group) **Group** *Project Name* **Repository: "***external-terminologies*"	NDS/DEM, (RDF Support)	
	Provide feedback	DCC	Can be an iterative process
	Incorporate feedback	NDS/DEM, (RDF Support)	
D	**De-identification rules**		
	Fill in the document highlighting the required de-identification rules using the template provided by the DCC.	NDS/DEM	
	Upload the filled template on project Git (Git group) **Group** *Project Name* **Repository: "***de-identification-rules*"	NDS/DEM	
	Provide feedback	DCC	Can be an iterative process
	Incorporate feedback	NDS/DEM	
	Approval process to be defined	DP	

2. Data Transfer Requests (DTRs)

Step	*Task*	*Responsibility*	*Comment*
1	Create the technical Data Transfer Requests in the BioMedIT Portal (Work instruction, one per data providing institution)	NDS/DEM	Attention: New process! The individual DTRs should not be approved by the DP until step 8 of this DTR workflow is completed
2	Evaluate the legal situation (is the DTUA signed?), approval of DTR	DCC (ELSI-Helpdesk)	
3	Prepare the following documents and upload them on the Git (Git group for all projects): – **Data transfer specification** (see template below) – **Cohort specification** *(e.g., use Excel template file from Project submission plus additional information if needed)* – **Reused SPHN concepts** *(csv extracted from the Confluence, include only concepts you want to request in this DTR, including additional information e.g., Scores or Lab tests, additional comments like e.g., mandatory for the project, specific subsets of code to be delivered)* – **Requested project concepts** *(csv; optional; include only concepts you want to request in this DTR, since your schema may contain concepts not relevant for all data providers, e.g., omics concepts or cohort concepts may not be requested from a UH)* **Folder structure and naming** **Group** *Project Name* **Repository**: "data-transfer" – **Folder**: "data-transfer-1" – **ReadMe** **Data transfer specification** – Technical DTR IDs from the BioMedIT Portal: – Status (in preparation/active/on hold/ stopped) – Frequency of data transfers: – Timeline: – Providers: – Cohort specification: *link to document on Git* – De-identification rules: *link to the document on Git* – RDF Schema: *link to schema on Git e.g., SPHN 2023.2 or LUCID 2023.1* – Reused SPHN concepts: *link to csv on Git* – Requested project concepts: *link to csv on Git* – Comments:	NDS/DEM	A common structure of this folder makes life easier for DP and DCC teams

	– **Subfolder: "***cohort-specification*"**Subfolder: "***reused-sphn-concepts*" – **Subfolder:** "*requested-project-concepts*" – **Folder**: "data-transfer-2" …		
4	Notify the DCC once the documents are final and ready for evaluation: Email to dcc@sib.swiss with reference [NDS/DEM Name and DTR Nr.], containing the link to the GitLab repository.	NDS/DEM	
5	Evaluate the documents	DCC	Can be an iterative process together with the NDS/DEM with consultancy of HIT-STAG if necessary
6	Final approval and notification of NDS/DEM, add tag "DCC-approved-dtr-<number>-<project>"	DCC	
7	Communicate to UHs that the current DTR is approved and documents can be accessed on the Git (DP Coordinator, UH-specific project coordinators, NDS/DEM PIs and DM, cc HIT-STAG representatives)	DCC	
8	Trigger of UH internal (governance) processes, if needed	DP	
9	Approval of the technical DTRs in the BioMedIT Portal by data providers (DP) indicating that DP are ready	DP	This should only be done once all of the above criteria/steps have been met
10	Start of the data transfers	DP	
11	In case a data transfer request is stopped or discontinued change the status in the README. Inform DCC by email dcc@sib.swiss *Each change in the specification results in a new Data transfer request on the Git.*	NDS/DEM	
12	Inform all involved DPs by email	NDS/DEM and DCC	

NDS/DEM=National Data Streams/Demonstrator Projects; DP=Data Provider; RDF Support=Data Management / RDF support on the BioMedIT nodes; DCC=Data Coordination Center, DTUA= Data Transfer and Use Agreement, DTR= Data Transfer Requests, UH= University Hospitals.

Example: A full example of a dummy project can be found on the Git (here)