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.

Version 4, 04th September 2024

1. Revision history

Version	Date	Changes
1	01/09/2023	*Initial version*
2	15/11/2023	– Added detailed process to section "D. De-identification rules"
3	25/04/2024	– Added revision history – Added RDF Data formats in the specification readme – Added section "E. Ethic documents" – Improved description of DTR step 11. – Updated links to external resources – Added the possibility to specify the RDF data format in the data specifications. – Updated DTR process (section 3) to reflect integration in BioMedIT Portal: o Removed the necessity of specifying the technical DTR IDs in the GitLab specifications.
4	04/09/2024	– Replaced DCC (FAIR Data Team) contact address

2. 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"
	  – 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*" (HTML for website documentation of the schema)
	– **Folder:** "version-2" …
		NDS/DEM, (RDF Support)	A strict adherence to the folder structure is necessary so that the SPHN Connector can directly access the 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: fair-data-team@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. For all rules where the project derivates from the default rules, provide a reason why this is needed.	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	
	In case any rules are deviating from the default, approval of hospitals is needed. DCC to contact UHs via HIT-STAG	DCC	
	HIT-STAG representatives are responsible to check hospital internally according to the hospital internal processes.	DP	
	In case hospitals do not agree with the de-identification rules, adapt accordingly	NDS/DEM	
E	**Ethics Documents**		
	Prepare the required ethics documents depending on the project setup (e.g., documents submitted to the ethic committees, ethics approvals, etc.). Make sure that the documents are signed by all the required parties.	NDS/DEM	
	Upload the documents on project Git (Git group) **Group** *Project Name* **Repository: "***ethics-documents***"**	NDS/DEM	

3. Data Transfer Requests (DTRs)

Step	Task	Respon-sibility	Comment

1	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
	  – 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*
	  – RDF data format: (.ttl/.trig/.nq)
	  – Comments:
	
	  – **Subfolder: "***cohort-specification***"**
	  – **Subfolder: "***reused-sphn-concepts***"**
	  – **Subfolder: "***requested-project-concepts***"**
	– **Folder**: "data-transfer-2"
	  …
		NDS/DEM	A common structure of this folder makes life easier for DP and DCC teams

2	Create the technical Data Transfer Requests in the BioMedIT Portal (Work instructions, one per data providing institution) Important: The link to the data specification needs to be provided during the technical DTR creation. Ensure that all your documents are thoroughly prepared as outlined in Step 1.	NDS/DEM	
3	Legal approval step on BioMedIT portal. Evaluate the legal situation (is the DTUA signed?).	DCC (ELSI-Helpdesk)	

4	Data specification approval step on BioMedIT portal. Utilize the link provided to access the data specifications in the project GitLab space.	DCC	Can be an iterative process together with the NDS/DEM with consultancy of HIT-STAG if necessary
5	Tagging of the data transfer repository: "DCC-approved-dtr-\<number>-\<project>" Replace the link to the data specification with the approved one in all related technical DTRs on BioMedIT Portal.	DCC	
6	Approval of the data specifications on the BioMedIT Portal.	DCC	
7	Evaluation of the DTR by the DP: – All technical measures are in place to send the data – There is an existing legal basis for the Data Transfer – This data delivery has been approved through internal governance	DP	
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	
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 fair-data-team@sib.swiss. *Please note that any change, no matter how minimal, to the approved DTR will result in discontinuing the current approved DTR, necessitating the creation of a new 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 GitLab projects space: Example project