

Foundation for Research and Technology - Hellas, Institute of Computer Science Information Systems Laboratory Centre for Cultural Informatics

X3ML Toolkit

Overview

The Centre for Cultural Informatics works as competence center for the CIDOC CRM (ISO 21127), by building up and exchanging application know-how, consultancy to implementers and researchers, and contribution to the dissemination, maintenance and evolution of the standard itself. CCI in collaboration with CIDOC for the last 20 years, received basic empirical input and developed a significant number of mapping specifications using a formalism based on the declaration of equivalences and the interpretation of each source schema as a set of nodes and links. Such mappings were used to validate the CRM. The X3ML mapping definition language, a declarative XML language, was the culmination of this experience and formalism. It describes schema mappings in such a way that they can be collaboratively created and discussed by experts who are able to produce the semantics of a mapping. It is human and machine readable and is the ultimate communication means on the semantic correctness of the mapping.

Following the X3ML definition, we developed and used an efficient transformation algorithm which processes the declarative X3ML statements and produces equivalent rdf statements. Different cases of semantic heterogeneity that were encountered in different applications are covered. In parallel, the CIDOC CRM SIG started working on the SYNERGY framework, a rich and comprehensive Reference Model for a better practice of data provisioning and aggregation processes, primarily in the cultural heritage sector, but also for e-science. The aggregation of heterogeneous data from different institutions has the potential to create rich data resources useful for a range of different purposes, from research to education and public interests. Along this line, CCI has developed the X3ML Toolkit, a set of small, open source, microservices that follow the SYNERGY Reference Model. They are designed with open interfaces and they can be easily customized and adapted to complex environments.



SYNERGY Schema Matching Flow Network



Target Domains

The X3ML Toolkit can be used for mapping and transforming data from cultural heritage institutions, such as archives, libraries, and museums, but equally for research institutes of descriptive sciences such as earth sciences, biodiversity, clinical studies and e-Health.

Description

The X3ML Toolkit consists of a set of software components that can be used independently, integrated in external environments or as a software tool that assist the data provisioning process: (i) the Syntax Normalizer is used to normalize the provider's records, exploiting local syntax rules and producing the Effective Provider Schema; (ii) the Mapping Memory Manager (3M) is a managing system suitable for handling the X3ML mapping definition files; (iii) the 3M Editor, is an application suite that helps the experts complete the Schema Matching Definition and produce the X3ML mapping definition file in a friendly, intuitive environment; (iv) the Source and the Target Schema Validator components assist the experts in selecting the valid paths with respect to the corresponding schemata, preventing mistakes, while the Source and the Target Schema Visualizer help the user navigate through all source and target elements; (v) the Instance Generator Rule Builder supports the specification of the instance generation rules that define the URI generation policy for each target class, complementing the Schema Matching Definition with the instance generation policies, producing the final X3ML Mapping Definition file; (vi) the matching process is also supported by the Mapping Suggester, which makes use of "mapping memories" of similar cases as they are collected and cached from the user community; (vii) the X3ML Engine component receives as input the X3ML Mapping Definition file and the normalized input from the provider and is responsible for transforming the records to the aggregator format, so that they can be ingested to the Aggregator Institution.

P & ISSUER_ID == PR_ID	P17_waa_motivated_by				300	Mapping : OEAW Coin	s DB					
	E7_Activity e [ia1]	[P2_has_type] [E55_Type = "Issuing"]										
	P14_camed_out_by			(Ten out is an excited at	Info	Matching Table Analysis	About					
R MSSUER	E39_Actor de			role.)	View mer							Mean mod
SOURCE	TARGET	CONSTANT EXPRESSION	IF RULE	COMMENTS () X	Colores .	0000004	TADOLTA	CONSTANT FROM	F C CICH	IF RULE	COMPANY	
D 🔲 //COIN	E22_Man-Made_Object	(PS2_hos_curre (E40_Legal_Body		11 C	Expand /	SOURCE	E22 Man-Made Object	[P52 has curren	(E40 Legal Body	ŕ.	COMMENTIN	
		nt_owner] = "OEAW"] 1/250 has currer [Ed3 Level Borts		A 1	Top				"OEAW"			
		nt_keeper] = "OEAW"]		× 1	Britten	//COIN			"OEAW"			
		[P2_has_type] [bb5_type = "coins"] [P46_forms_par [E78_Collection		()					[E55_Type = "coins"]			
		Lot "dFMRO"			ANL				"dEMRO"			
P Source Relation	Target Relation		Add Bule v	Add Commentation die	P	1 10	Pt_is_identified_by					_
DENOMINATION H +	↓ Р2_лаз_рре ж +				R	D 10	E41_Appellation				(2% map local identifie	
Add Intermediate	Add Intermediate						P108_was_produced_by					
					P	COUNTRY_ID == COUNTRY_ID	E12_Production e [p1]		[E55_Type = "Roman coloa"]		(E12 Production [x1]->	PC1
R Source Node		Relation ×	Add Rule -	Example Target			1 Pt0_falls_within					
Langeoupen. x +	Coolinearanticult R + K	→ Р2_hts_0pe ж ч		tor example .xur Denc	R	//COUNTRY	E4_Period	[P2_has_type]	[E55_Type = "country"]			
	core#Concept H * X	Entity		Add Comment about +			P30_custody_transferred_through					
	Add additional class	ESS_Type × *			P	FIND_SPOT_ID == FS_ID	E10_Transfer_of_Custody e [tc1]					
	Add instance info +	Constant			0	COLUMN SPOT	P7_fooe_prace_at					
		: denomination			R.	inno_aron	L CSJ Place					_
		Add instance info •			р	FIND MANNER ID == FM ID	E10 Transfer of Custody				Coin federa averal ma	v have by 1
		Enery					↓ P2 has type					
		Constant			0	T //FIND MANNED	E55_Type de					
		Constant					core#Concept					
		1 ORIGINATION					P30_custody_transferred_through F10_Transferred_floateckbe11					
		Add instance into +			F	P 1 FIND DATE	1 P4 has time-soan					
		Add Constant Expression					E52 Time-Span e [ts1]					
DENOMINATION == DEN_ID							P81_ongoing_throughout					
1 DEN METAL == MET ID	P45_consists_or				R	FIND_DATE	rdf-schema#Literal de					
	E57 Material						P108_was_produced_by					
	core#Concept			1-3	P	AUTHORITY_ID == AUTH_ID	E12_Production					
	P108_was_produced_by				R	I //AUTHORITY	Fill Pariod de	[F2 has type]	TESS Type = "country"			
A MANUJO MINIJID	L12_Production e [p1]						1 P108 was produced by	1				_
A COLUMNT	Fill Legal Body @ [b1]	[P2 has hoe] [E55 Type = "mint"]					E12 Production e [p1]					
	1 P108 was produced by	the state of the s			P	ISSUER_ID == PR_ID	P17_was_motivated_by					
	E12_Production e [p1]						ET_Activity e [ia1]	[P2_has_type]	[E55_Type = "Issuing"]			
							Th - 20					
The 3M Editor						i ne 3M Viewer						

Additional Information

The X3ML toolkit has been partially supported by the projects ARIADNE, PARTHENOS, BlueBRIDGE, VRE4EIC and is provided under open source license. More information at: http://www.ics.forth.gr/isl/X3MLToolkit 3M is accessible at http://www.ics.forth.gr/isl/3M



3M

Contact details:

Maria Theodoridou maria@ics.forth.gr www.ics.forth.gr/isl/cci.html

FORTH-ICS