

Notes

Session 1: The SSHOC Reference Ontology (SSHOCro): Modeling the SSHOC data life cycle

Keynote speaker: Athina Kritsotaki, FORTH

The SSHOC Reference Ontology is built as an extension of CIDOC CRM and it is a workflow model. It is an event-based ontology that aims to organize knowledge and information distributed across various primary sources of information in the Social Sciences and Humanities Open Cloud.

An example from Social Science Research instantiating the SSHOCro

The example is drawn from material published through the Open Science Framework (OSF) in the context of the Social Sciences Replication Project (SSRP).^[1] In 2016, a group of researchers undertook the task of replicating 21 experimental studies that had been published in *Nature* and *Science* magazines, between 2010 and 2015, to determine the robustness of the reported results. These replications were conducted in the context of the Social Sciences Replication Project,^[2] initiated by the Center for Open Science.^[3]

The goal of the SSRP was to replicate experimental studies with as minimal deviations as possible from the original experimental designs, procedures and methods followed in all the stages of the original studies. Replicating experimental studies allows one to assess the robustness of results arrived at through experimental observation. Given that successful replications increase the trust in scientific findings, they are considered pivotal, especially when they involve influential studies that predate the development of new methods, approaches and/or data. Multiple research groups were engaged in the Social Sciences Replication Project, each assigned to replicate a set of experiments, which, in their turn, included one or two distinct stages of **(i) data collection, (ii) data preparation, and (iii) data**



interpretation. Whether one or two rounds of replications took place depended on the success or failure of the experiment to replicate on the first round. The analysis stopped after the second round, irrespective successfully replicating the results obtained by the original experiment.

The study chosen to exemplify the classes and properties of SSHOCro, was the replication of *Balafoutas & Sutter (2012)*, the hypothesis of which was that preferential treatment increases women's competitiveness, without reducing men's in a lab environment. The results extrapolate in policymaking.

The researchers who performed the replicating study (F. Holzmeister, J. Huber, M. Kirchler, J. Rose) followed the experimental design of the original study, used the same tools for data collection, analysed them using the same tests and packages and arrived at, essentially, the same conclusions. The data, the procedure and the software for collecting it, the packages and the scripts used for data analysis are available for the replication through the OSF.^[4] However, they are not available for the original study.

In what follows, the reader can navigate through diagrams that document the information relayed by the example. **Figure 1** below represents the *Replication of Balafoutas and Sutter (2012)* at a macro-level. The *Replication of Balafoutas and Sutter (2012)* instantiates a Project Activity, which forms part of a larger Project Activity, undertaken by the group of researchers from the Innsbruck University division of the SSRP community. The SSRP, in its turn, is the overall Activity Project in the context of which the aforesaid Project Activities form a part of. The SSRP as a whole *offered a service*, namely the creation of a repository for replication studies in the Social Sciences, comprising materials (i.e. datasets plus scripts and software to operate on them) as well as reports and/or other publications. This transitive parthood relation is represented in the diagram by *the P9i forms part of* relations among the different sub-projects and the overall one. The instance of SHE Project Activity, *SSRP_Replication of Balafoutas and Sutter (2012)*, can be further broken down to the series of discrete sets of activities, each corresponding to a different stage in the research (collection, preparation-connection and interpretation).



Figure 1: Replication of Balafoutas and Sutter (2012) at a macro-level

Replication studies depend –to a great extent –on the scientific methods used in the studies the results of which they put to the test. That dependence is manifold and relates to an effort for as little deviation as possible from the research protocols observed in the original studies, on behalf of the scholars conducting the replication. The processes and methods undertaken during data collection, data preparation-connection and data interpretation in the replicate study are bound by the ones observed in the original. Replications are data driven and it is important that the relation among any data produced and manipulated in the context of the replication and the original study be truly comparable. In that sense, one must be able to express how they relate to one another, which is what **Figure 2**, below, aims for –namely to

express the relation of datasets to the project activities that produced or deployed them in the course of a research project plus the relation between datasets across different projects.

In particular, this is the case with *SHE Dataset –SHR incorporates: SHE Dataset* (instantiated by the value pair “*Datasets created and analyzed in the context of Balafoutas & Sutter 2012*” and “*Balafoutas & Sutter (2012) -replicate data*”, below).

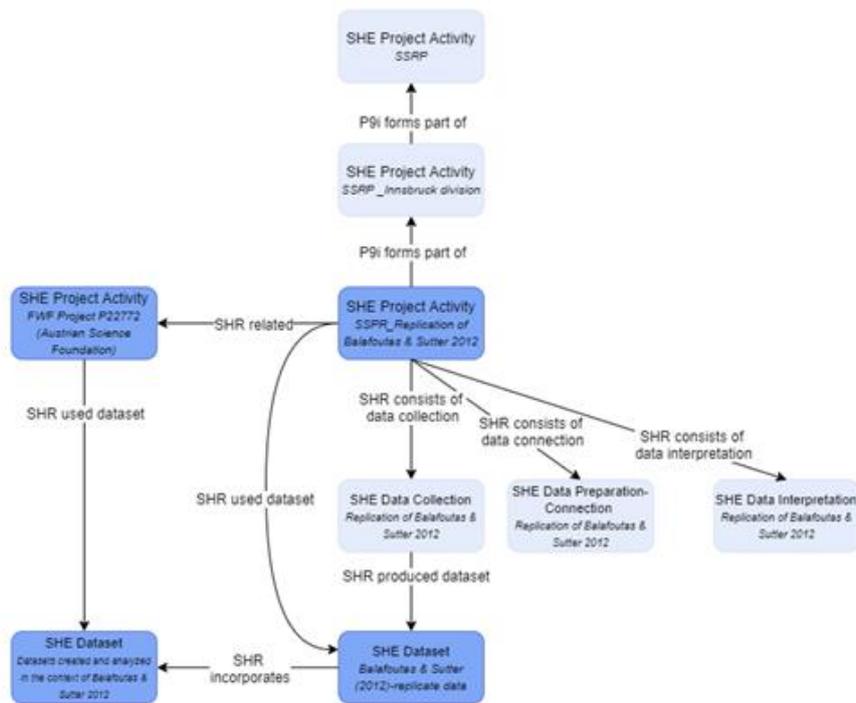


Figure 2: Data reuse in SSRP subprojects

Figures 3 through 5 illustrate the sets of activities undertaken during each of the distinct phases of the Project Activity *SSRP_Replication of Balafoutas and Sutter (2012)*. In particular, **Figure 3** illustrates the data collection process, **Figure 4** the stage of data preparation and connection, **Figure 5** the process of interpreting the collected data (post cleaning);

The data collection activity, linked to the time that it occurred –the place not being specified – to the actors that participated in it –and the capacity with which they participated –the software used for collecting the data, the purpose the data collection served and its type – producing vs. finding and/or reusing.

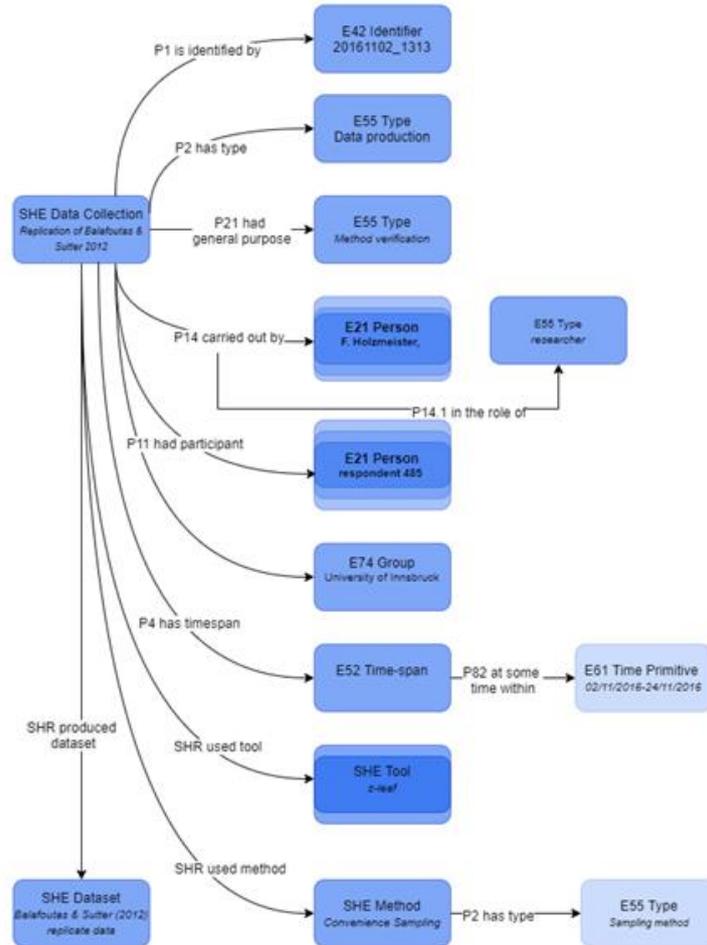


Figure 3: The data collection process

Figure 4 represents the data preparation-connection stage; it comprises a data preparation activity, linked to the time that it occurred –the place not being specified –to the actors that participated in it –and the capacity with which they participated –the software used for connecting the data, the methods deployed and their type. SHE Publication appears in the diagram above despite not being instantiated in the example, to demonstrate there can be publishing activities at all stages of the research project.

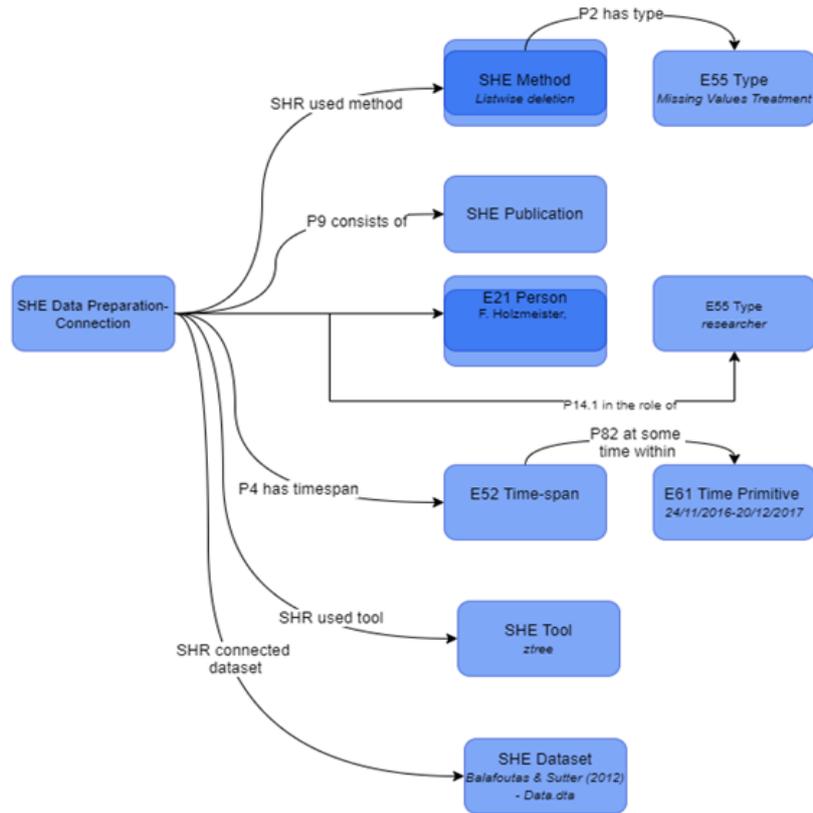


Figure 4: The data preparation-connection process (preprocessing phase)

Figure 5 represents the data interpretation; it comprises a data preparation activity, linked to the thing it was an interpretation of, the time during which it occurred, the actors that participated in it –and the capacity with which they participated –and the methods deployed for analyzing data and the software used in the analysis. The interpretation stage is further analyzed to its constituent parts –namely a set of activities of publishing data and replication

reports/papers. Publishing activities are represented in greater detail in the following

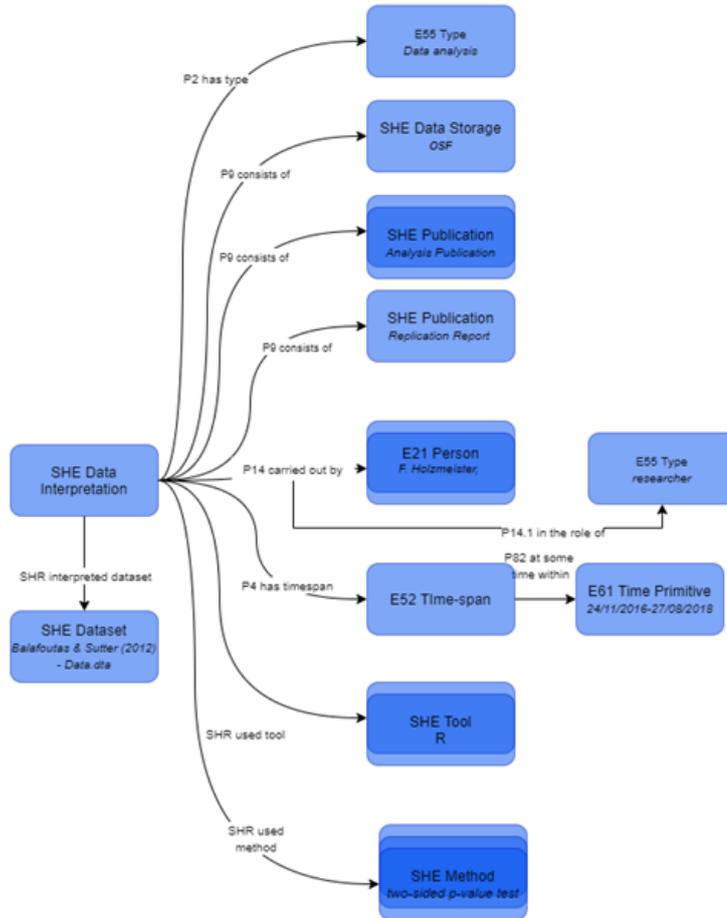


diagram.

Figure 5: The data interpretation process

[1] For more information on the project –in terms of overall goals, methods used, institutions and research teams involved, evaluations etc. –the reader is referred to Camerer C.F. et al. (2018). *Evaluating the replicability of social science experiments in Nature and Science between 2010 and 2015*. Open Science Framework. Retrieved February 25, 2020, from <https://osf.io/pfdyw/>

[2] All materials for the Social Sciences Replication Project (2016) are made available via the corresponding Open Science Framework (OSF) directory: <https://cos.io/our-services/research/ssrp-overview/>

[3] The designated website of the Center for Open Science is: <https://cos.io/>

[4] Holzmeister F., Huber J., Kirchler M., Rose J. 2018. *Replication of Balafoutas and Sutter (2012)*. Open Science Framework. Retrieved February 25, 2020, from <https://osf.io/m8qav/>