

What Data Get to Travel in High Energy Physics? The Construction of Data at the Large Hadron Collider



Koray Karaca

Abstract In present-day high-energy physics experiments, experimenters need to make various judgments in order to design automated data processing systems within the existing technical limitations. In this chapter, as a case study, I consider the automated data acquisition system used in the ATLAS experiment at the Large Hadron Collider (LHC) located at CERN, where the Higgs boson was discovered in 2012. I show that the design of this system involves both theoretical and experimental judgments each of which has different functions in the initial data journey through which usable data are constructed out of collision events detected by the ATLAS detector. I also explore what requirements the foregoing judgments impose on the LHC data in terms of usability, mobility and mutability. I argue that in present-day HEP experiments these aspects of data are distinct but related to each other due to the fact that they are subjected to some common requirements imposed by the theoretical and experimental judgments involved in the design of data acquisition systems.

1 Introduction

The introduction of computer technologies to experimental high-energy physics (HEP) experiments in the fifties and sixties resulted in the automation of data processing in HEP experiments (Galison 1997). Continuous advances in computer technologies have led to the ever-increasing automation of data processing in experimental HEP. This has made it possible to process increasingly large and complex data produced by increasingly more advanced particle detectors and colliders. As a result, experimental HEP has been progressively data intensive over the past 60 years, and this has been accompanied by important changes not only in terms of methods, techniques, and tools employed in HEP experiments (Franklin 2013; Gutsche et al. 2017), but also in terms of organizational structures (Boisot et al. 2011; Knorr-Cetina 1999) and authorship (Galison 2003) in experimental HEP collaborations.

K. Karaca (✉)

Department of Philosophy, University of Twente, Enschede, The Netherlands

e-mail: k.karaca@utwente.nl

© The Author(s) 2020

S. Leonelli, N. Tempini (eds.), *Data Journeys in the Sciences*,
https://doi.org/10.1007/978-3-030-37177-7_3

The ATLAS and CMS experiments¹ currently running at the Large Hadron Collider (LHC) located at CERN represent the state of the art in automated data processing in HEP experiments, as the level of automation achieved in these experiments is unparalleled in previous HEP experiments. While automation enables processing unprecedentedly large and complex data in the foregoing LHC experiments, it greatly reduces the need for human intervention in data processing. However, automation does not diminish the role of human judgments in this process. As I will discuss in this chapter, experimenters at the LHC need to make various judgments to be able to design automated data processing systems within the existing technical limitations.² As a case study, I will examine the automated data acquisition system used in the ATLAS experiment. I will argue that the design of this system involves both theoretical and experimental judgments each of which has different functions in the automation of data processing in the ATLAS experiment. I will also explore what kinds of requirements the foregoing judgments impose on the LHC data in terms of usability, mobility and mutability, which are the general aspects of data in physical and biological sciences (Leonelli 2016).

In addressing the foregoing issues, I shall make use of the notion of *data journey*, which is a useful metaphor to characterize various processes that data undergo in experiments performed in physical and biological sciences (ibid.). In these experiments, data journeys start with the process of data acquisition. Some of the philosophical aspects of this process have already been discussed in the context of the LHC experiments (see, e.g., Morrison 2015; Beauchemin 2018; Karaca 2017, 2018), and also in other contexts in this volume. In a case study concerning ocean science, Gregor Halfmann (in this volume) discusses the initial stage of data acquisition where data is first produced. In a case study concerning astronomy, Götz Hoeppe (in this volume) discusses aspects of data acquisition concerning data interpretation. In this chapter, I will focus on the initial data journey in the ATLAS experiment that links the production of collision events at the LHC to the stage of data acquisition where *usable* data are constructed out of collision *events* detected by the LHC, prior to the stage of data analysis and modeling (Karaca 2018; Leonelli 2019; Boumans and Leonelli in this volume).

In scientific experimentation, data usability means the fitness of experimental data for its intended uses, namely data analysis and data modelling, which are aimed at serving the objectives of an experiment. In the context of present-day HEP experiments, the term *data* is used to refer to *collision events* produced by collider systems such as the LHC and detected by detector systems such as the ATLAS and CMS detectors. In the terminology of HEP, the term event denotes “the record of all the products from a given bunch crossing,” (Ellis 2010, 6) which occurs when two beams of particles collide with each other inside the collider. In the ATLAS experiment, proton

¹The names of these HEP experiments are derived from the ATLAS (A Toroidal LHC Apparatus) and CMS (Compact Muon Solenoid) detectors located at the LHC.

²The details of the design of the automated data processing systems used in the ATLAS and CMS experiments are explained in the technical design reports of these experiments, see ATLAS Collaboration 2003; CMS Collaboration 2002.

bunches, rather than individual protons, collide inside the LHC at a rate of approximately 40 million times per second. These recorded collision events, amounting to petabytes ($=10^{25}$ bytes) of data, are then processed and finally digitally recorded on tapes in databases at CERN. I shall call the foregoing journey of the LHC data the *local data journey*, as opposed to the *global journey* that I take to refer to the journey of the LHC data concerning its dissemination to researchers located inside and outside CERN.

The plan of this chapter is as follows. In Sect. 2, I will discuss how the criteria for usable data are specified in the ATLAS experiment. Also, I will characterize the experimental strategy used to search for usable data in this experiment. In Sect. 3, I will examine the local data journey at the LHC and show how usable LHC data are constructed out of event fragments detected by the ATLAS detector. In the final section, I will argue that in the ATLAS experiment data mutability is required for data usability, and that the former is enabled by data mobility through the local data journey at the LHC. Furthermore, I will identify the judgments involved in the design of the ATLAS data acquisition system. I will argue that as a result of the requirements imposed by the foregoing judgments, usability, mutability, and mobility are related, though distinct, aspects of the LHC data during its local journey.

2 Selection Criteria and Search Strategy for Usable Data in the ATLAS Experiment

The ATLAS experiment at the LHC is a multi-purpose HEP experiment with two sets of objectives (ATLAS Collaboration 2003, Sect. 4): (1) to test the predictions of the present models of HEP concerning new particles, including the Higgs boson predicted by the Standard Model (SM)³ of elementary particle physics and the particles, such as new heavy gauge bosons, superpartners and gravitons, predicted by the theoretical *models beyond the SM* (BSM models) that have been offered as possible extensions of the SM model, such as super-symmetric and extra-dimensional models (Ellis 2012); and (2) to search for *unforeseen* physics processes, i.e., those that have not been predicted by the present HEP models, including possible deviations from the SM at *low* energies. As I shall show in this section, the diversity of the objectives of the ATLAS experiment has a crucial bearing on what is considered usable data in this experiment, and also on the procedure through which this data is acquired.

The first set of objectives of the ATLAS experiment concerns a range of predictions concerning different kinds of heavy particles (including the SM Higgs boson) that are predicted to be produced at high energies, while its second set of objectives concerns unforeseen physics processes which might occur at both high and low ener-

³The SM consists of two different gauge theories; namely, the electroweak theory of the weak and electromagnetic interactions, and the theory of quantum chromo-dynamics which describes the strong interaction.

gies. This means that the collision events relevant to the first set of objectives of the ATLAS experiment are also relevant to its second set of objectives concerning the discovery of unforeseen physics processes that might occur at high energies. Therefore, the objectives of the ATLAS experiment require different, but partly overlapping, types of collision events to be acquired during the stage of data acquisition.

In the context of present-day HEP experiments, collision events that have the potential to serve the objectives of the experiment are often referred to as *interesting events*. In the case of the ATLAS experiment, the *signatures*⁴ predicted by the SM for the Higgs boson are high transverse-momentum (p_T)⁵ photons and leptons,⁶ and the ones predicted by the BSM models for new particles beyond the SM, such as new heavy gauge bosons W' and Z' and supersymmetric particles, are high p_T single particles, namely photons and leptons, high p_T jets as well as high missing and total transverse energy (E_T).⁷ The aforementioned high p_T and E_T types of signatures might be produced at the LHC as a result of the decay processes involving the Higgs boson and the aforementioned particles predicted by the BSM models. The same types of signatures might also be produced at the LHC as a result of some *unforeseen* physics processes occurring at high energies (i.e. approximately above 10 GeV). This means that the collision events containing high p_T and E_T types of signatures are relevant to both sets of objectives of the ATLAS experiment, thus making them *interesting* for the process of data selection.⁸ For this reason, in the ATLAS experiment, the selection of the interesting events relevant to the predictions of the SM and BSM models, as well as to the discovery of unforeseen processes at high energies, is performed by using selection criteria that consist of *only* the aforementioned high p_T and E_T types of signatures. These selection criteria are often referred to as *inclusive triggers*, in the sense that they constitute the main set of selection criteria in the trigger menu used in the ATLAS experiment.

As the above discussion indicates, the range of interesting events in the ATLAS experiment includes a wide variety of high p_T and E_T types of signatures across a wide range of p_T and E_T values, i.e., approximately from 10 GeV to 1 TeV. The technical limitations in terms of data storage capacity and data process time make it necessary to apply data selection criteria to collisions events themselves in *real-time*, i.e., during the course of particle collisions at the collider (ATLAS Collaboration

⁴The term *signature* is used in experimental HEP to denote stable sub-atomic particles or energies into which unstable sub-atomic particles decay as a result of a physical process.

⁵Transverse-momentum is the component of the momentum of a particle that is transverse to the proton-proton collision axis, and transverse-energy is obtained from energy measurements in the calorimeter detector.

⁶A lepton is a spin $\frac{1}{2}$ particle that interacts through electromagnetic and weak interactions, but not through strong interaction. In the SM, leptons include electron, muon and tau, and their respective neutrinos.

⁷In this context, the term *high* refers to the p_T and E_T values that are approximately of the order of 10 GeV for particles, and 100 GeV for jets.

⁸The foregoing types of signatures also differ among each other, as the predictions to which they are relevant, namely those by the SM and the BSM models, are different from each other (for details, see ATLAS Collaboration 2003, Sect. 4; Karaca 2017).

2012). Moreover, due to the aforementioned technological limitations, only a minute fraction of the interesting events could be selected for further evaluation at the stage of data analysis. This necessitates, for the fulfillment of the objectives of the ATLAS experiment, that the trigger menu (i.e. the full list of data selection criteria) be sensitive enough to select the range of types of interesting events that will serve the entire range of objectives of the ATLAS experiment. If the trigger menu were not appropriate to this end, then the data selection procedure would be biased against certain types of interesting events. As a result, the ATLAS experiment would fail to achieve some of its objectives, as the fulfillment of a particular objective of the ATLAS experiment requires the acquisition of certain types of interesting events.

A major challenge in the ATLAS experiment is to perform data selection in an unbiased manner with respect to the various objectives of the experiment. This challenge has been addressed through a particular data selection strategy that aims at increasing the sensitivity of the trigger menu, and thus of the selection procedure. To this end, the foregoing selection strategy requires the trigger menu to be sufficiently diversified in terms of types of selection signatures that are appropriate for the various objectives of the experiment. Since the ATLAS experiment is largely aimed to test the SM's prediction of the Higgs boson and the predictions of the BSM models, the adopted strategy in the first place requires the trigger menu to be sufficiently diversified in terms selection signatures composed of only high p_T and E_T types of signatures relevant to the aforementioned predictions. This aims at extending the range of the relevant LHC data that could be acquired through the trigger menu.

In the ATLAS experiment, unforeseen physics processes might also occur at low energies, i.e., approximately below 10 GeV . Inclusive triggers are not appropriate for the search for novel p_T and E_T processes at low energies, as these selection criteria consist of only *high* p_T and E_T types of signatures. Therefore, the selection strategy adopted in the ATLAS experiment also requires the trigger menu to be sufficiently diversified in terms low p_T and E_T types of selection signatures. These selection signatures are referred as to *prescaled triggers* and determined by prescaling inclusive triggers with lower p_T and E_T thresholds ($<10\text{ GeV}$) (for details, see ATLAS Collaboration 2003, Sect. 4.4.2). In this context, *prescaling* means that the amount of events that a trigger could accept is suppressed by what is called a *prescale factor* in order for the selection process not to be *swamped* by the events containing vastly abundant low p_T and E_T types of signatures, so that the aforementioned first set of objectives of the ATLAS experiment is not endangered. Prescaled triggers are necessary for the trigger menu, and thus of the selection procedure, to be sensitive enough to the search for novel p_T and E_T processes at low energies. Since the events containing low p_T and E_T types of signatures have the potential to be of use for some SM studies of strong interactions (see, e.g., ATLAS Collaboration 2016) as well as to provide support for new physics searches at low energies, prescaled triggers are especially aimed at further extending the range of the LHC data relevant to the second set of the objectives of the ATLAS experiment.⁹

⁹Note that these events are also used to determine trigger efficiencies and detector performance.

3 Local Data Journey at the LHC

In the ATLAS experiment, the trigger menu is applied to collision events at three different levels through the use of what are called *trigger systems* (Ellis 2010).¹⁰ These are automated systems designed and used to select the desired events from the collision events. The first stage of the data selection process is carried out by the level-1 trigger system that provides a crude selection of the interesting events in real-time. In the ATLAS experiment, the initial event rate of the proton-proton collisions is ~ 40 MHz, corresponding to approximately 40,000,000 collision events per second. The first level of the data selection process is performed by the level-1 trigger system, whose technical features allow for an event-acceptance rate of 75–100 kHz. The second and third levels of the data selection process are respectively carried out by the level-2 and level-3 trigger systems, which are jointly called the *High-level Trigger and Data Acquisition System* (HLT/DAQ). Unlike the level-1 trigger system, which is hardware-based, the HLT/DAQ system is software-based, meaning that the level-1 and level-2 selection processes are performed directly by the specialized software algorithms according to the trigger menu. The level-2 and level-3 trigger systems have much smaller event-acceptance rates, which are respectively around ~ 2 kHz and ~ 200 Hz, and thereby provide finer selections of the desired events.¹¹ Therefore, in the ATLAS experiment, the initial event rate is gradually lowered from 40 MHz down to around 200 Hz at the end of the level-3 selection process, meaning that the interesting events are selected from the collision events at a ratio of approximately 200/40,000,000, i.e., 5 in every 1 million events.

The first stage of the data acquisition process is carried out by the level-1 trigger system that performs a *crude* selection of potentially interesting events from the collision events detected by the calorimeter and muon detectors, which are the components of the ATLAS detector system.¹² The level-1 trigger system produces a trigger decision within $2.5 \mu\text{s}$ and thereby reduces the LHC event-rate frequency of 40 MHz down to the range of 75–100 kHz. In addition to the calorimeter and muon detectors, the tracking detectors are also used in the ATLAS experiment.¹³ Since the event rate is so high and thus the trigger decision time is so short, it is technologically impossible for the tracking detectors to determine particle tracks quickly enough for the level-1 event selection. Only the *hit points* produced by particles inside the tracking detectors could be recorded. These space points are later assem-

¹⁰The treatment in this section is based on the ATLAS Technical Design Report (ATLAS Collaboration 2003), which is a technical document that contains the design information concerning the principal components and functions of the ATLAS data acquisition system.

¹¹Note that the aforementioned event-acceptance rates are valid only for the early data-taking run (Run-1) and have changed significantly during Run-1 and also during Run-2.

¹²ATLAS is a detector system that consists of different individual detectors, including the inner detector and the calorimeter and muon detectors.

¹³In HEP experiments, the tracking detectors are used to determine particle tracks as well as to measure the momenta of electrically charged particles by means of the curvatures of their tracks in a magnetic field.

bled by software algorithms in order to determine particle tracks. As a result, the data from the tracking detectors are not used directly by the level-1 trigger system for event selection. Moreover, due to the shortness of the level-1 trigger-decision time, even though the hit points are recorded, they are not completely read out from the tracking detectors during the level-1 selection. This means that the information (i.e., in terms of location in the detector, and p_T or E_T for each particle or jet contained, or associated missing E_T) necessary to fully specify a selected event is fragmented across the individual detectors of the ATLAS detector system, and that all pieces of this fragmented information are not assembled yet. Therefore, the full description of the event is not yet known, and as a result, the level-1 event selection is performed without *full granularity*, i.e., without the availability of data from all the channels of the individual detectors.

As shown in Fig. 1, the level-2 event selection begins when the sub-unit called Level-2 Supervisor sends (arrow 1)¹⁴ the results of the level-1 selection to the sub-unit called Level-2 Processing Unit (arrow 2). Unlike the level-1 trigger system, the level-2 trigger system uses the RoI data¹⁵ processed by the sub-unit called Read-out System (ROS) from all the sub-detectors of the ATLAS detector with full granularity. The event fragments, which are temporarily stored in the ROS, are accepted to the level-2 selection in small amounts. This way of performing event selection is called the *seeding mechanism* (ATLAS Collaboration 2003, Sect. 9.5.3.1). The ROS sends (arrows 2.1 and 2.2) to Level2Processing a subset of the event-fragments data, namely, the information regarding the locations (in the detector), momenta, and energies of the events selected at the level-1 selection. LVL2Processing sends (arrow 3.1) the information regarding the events accepted by the level-2 trigger system back to the ROS. LVL2Processing also sends (arrow 3.2) this information to LVL2Supervisor. LVL2Supervisor forwards (arrow 4) the same information to the sub-unit called Event Builder, which receives from the ROS the event-fragments data for the events selected by LVL2Processing. Event Builder (arrow 5.1) requests from the ROS the event-fragments data for the events selected by the LVL2Processing unit. Upon this, ROS (arrow 5.2) sends the event fragments to the Event Builder. The component called Sub-Farm Input (SFI) of the Event Builder assembles the event fragments associated with each selected event into a single record. At this stage, the full description of each selected event is available. The events that have been built are then passed (arrow 6) to the sub-unit called Event Filter Processor (EFP), through which the level-3 event selection, which is also called “event filter” (EF) selection, is carried out by specialized software algorithms (arrow 7).¹⁶ The events that have passed the level-3 selection are then sent (arrow 8) to the sub-unit called Sub-Farm Output (SFO) for permanent storage and offline data analysis.

¹⁴Arrows refer to Fig. 1.

¹⁵The regions in the ATLAS detector that contain signals for interesting events are called *regions of interest* (RoIs). The RoIs and the energy information associated with the signals detected in the RoIs are together called the RoI data.

¹⁶Note that in Fig. 1, the correct arrow numbers for the messages “EFSelection” and “SendEvent” should be “7” and “8” respectively.

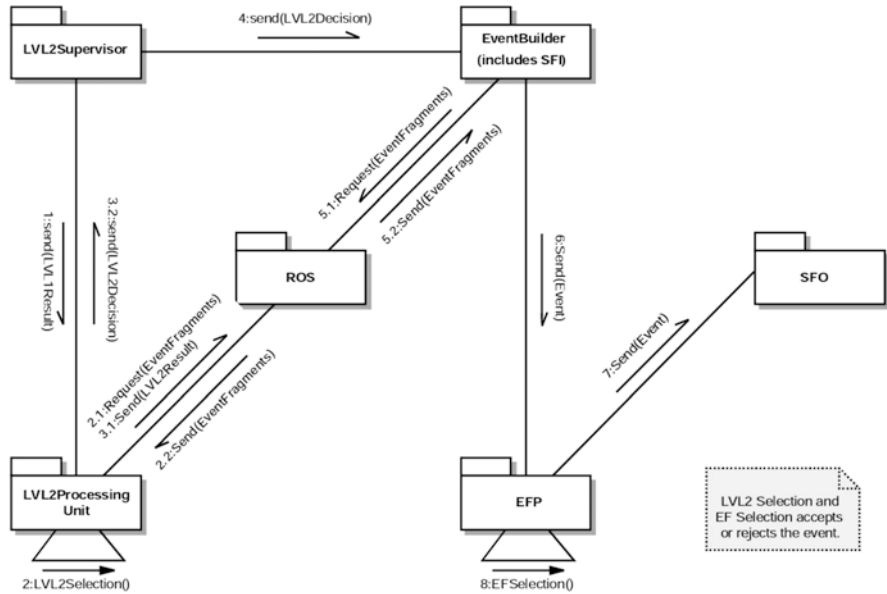


Fig. 1 The relationships between the different components of the HLT/DAQ system in the ATLAS experiment. (Source: Fig. 9–2 in ATLAS Collaboration 2003)

The details of the level-2 and level-3 selection processes are not shown in Fig. 1. These selection processes are carried out by the Event Selection Software (ESS) system, which is a software component of the HLT system (ATLAS Collaboration 2003, Sect. 9).¹⁷ The level-2 selection of an event is carried out in a series of *steps* each of which consists of two stages. In the first stage, the event is partially reconstructed, meaning that the trigger elements (TEs)¹⁸ associated with the event are refined and reconstructed by the *reconstruction algorithms* according to what is called the *sequence table* of the step. Each sequence in this table consists of an input TE and a reconstruction algorithm that is to be executed to refine and reconstruct an input TE into an output TE.¹⁹ In the second stage, the event partially reconstructed undergoes a selection process based on what is called the *menu table* of the step that contains a list of the selection signatures required for this step.

The Step Handler initiates the first stage of the level-2 selection by executing the Step Sequencer to access the list of the *active* input TEs associated with an event

¹⁷For future reference, note that the following units to be mentioned in what follows, namely, Step Handler, Step Sequencer, Step Decision, Step Controller and Result Builder, are the software components of the ESS system that steers the HLT selection process.

¹⁸A TE denotes one specific signature identified by the level-1 trigger system, e.g., “*e25 \bar{t}* ”. A TE is said to be *active* if it has previously satisfied a selection signature at the level-1 selection, or at the previous step of the level-2 selection, if the step under consideration is not the first step of the level-2 selection.

¹⁹Reconstruction algorithms are a class of HLT algorithms that act on the RoI data with full granularity from all sub-detectors to find new features associated with input TEs, such as a track or an isolation requirement.

selected by the level-1 trigger system. The Step Sequencer next compares the list of the active TEs with the required TEs given in the sequence table of the step. For all matching TEs, the Step Sequencer executes the reconstruction algorithms to refine and reconstruct the input TEs into the *output* TEs according to the sequence table of the step. The Step Sequencer also creates the list of the output TEs for the implementation of the seeding mechanism discussed earlier. The Step Sequencer also marks each output TE as “seeded by input TE” depending on from which input TE it has been previously created. Then, it passes each output TE to the relevant *hypothesis algorithms*—another class of HLT algorithms—that decide whether the TE is valid, depending on whether its reconstructed features are consistent with its physics interpretation. For example, if a track or an isolation requirement associated with a TE is found by a reconstruction algorithm, then the relevant hypothesis algorithm determines whether this track or isolation requirement matches the physics interpretation of the TE. The hypothesis algorithms *activate* the validated TEs and discard the invalidated TEs by deactivating them.

The Step Handler initiates the second stage of the level-2 selection by calling the Step Decision to access the list of the active output TEs, i.e., the TEs validated by the hypothesis algorithms in the first stage of the level-1 selection. The Step Decision compares the list of the active output TEs with the required selection signatures given in the menu table of the step. For the TE combinations that match the selection signatures in the menu table, the Step Decision creates a list of the satisfied signatures that consist of those matching TE combinations. The event is accepted for the next step by the Step Decision, if the TE combinations it contains satisfy at least one signature given in the menu table of the step; otherwise it is rejected and thus not considered for the level-3 selection. The Step Decision sends the information regarding the decision about the event to the Step Handler that will initiate the next step configured with a different sequence table and a menu table. The level-2 selection of an event ends at the step where it is rejected, or it continues until all required steps are completed, indicating that the event is finally accepted for the level-3 selection.

If an event is accepted at the level-2 selection, the Step Controller executes the Result Builder to provide the information necessary to *seed* the level-3 selection. This includes all satisfied signatures and the associated TE combinations, as well as the level-1 RoI data. The Result Builder assembles all these data-fragments, and the results are subsequently used for the seeding of the level-3 selection. The level-3 selection is implemented and coordinated by the Step Handler in the similar way as the level-2 selection is carried out as described above. But, the level-3 selection differs from the level-2 selection in that the TEs are now the active TEs of the level-2 selection, and that more sophisticated HLT algorithms are used to achieve a much finer event selection. As has been mentioned previously, the events that have passed the level-3 selection are stored in the Sub-Farm Output for data analysis. This marks the end of the local journey of the LHC data.

The collision events that have been rejected by the level-1 and level-2 trigger systems are removed from the data selection system. However, all the data selection operations carried out by the ATLAS data acquisition system are recorded by the system called Online Bookkeeper that produces logs stored in the form of logbook data (ATLAS Collaboration 2003, Sect. 10.4.1.2). Therefore, the ATLAS data

acquisition system is traceable in the sense that the decision regarding the acceptance or rejection of an event (already selected by the level-1 trigger system) by the level-1 and level-2 system systems can be reassessed by using the logbook data.

The LHC data is disseminated to the researchers located outside CERN through its global journey implemented by the ATLAS Distributed Data Management system (ADDMM) where the acquired collision events are digitally written to *datafiles* aggregated into what are called *datasets* (for details, see Branco et al. 2008). The latter are disseminated through its four-tier hierarchical structure.²⁰ Tier-0 is the CERN Data Center where datasets are created, stored and distributed to Tier-1 which consist of (currently) 13 computer centers located in the following countries: Canada, Germany, Spain, France, Italy, Nordic countries, Netherlands, Republic of Korea, Russian Federation, Taipei, UK, and US. Tier-1 temporarily store datasets and distribute them to Tier-2 which consists of computer centers located typically at universities and similar scientific institutions. There are currently 150 Tier-2 sites around the world. Researchers located outside CERN can access data sets (for the purpose of data analysis) through Tier-3 which consists of local computer clusters located at universities and similar research centers or even through individual personal computers.

4 Conclusions

The technical limitations at CERN in terms of data storage capacity and data process time do not allow applying the trigger menu to the detected events without subjecting them to the construction and selection processes that make up the local data journey in the ATLAS experiment. Since the requirements for data usability are specified by the selection criteria in the trigger menu, data mobility is necessary for data usability and constitutes an essential aspect of the ATLAS data acquisition process. During the local data journey, collision events detected by the ATLAS detector system are constructed out of the fragments of proton-proton collision events that are produced by the LHC and detected by the ATLAS detector system. The first part of the local journey is a construction process in the sense that event fragments are assembled by the level-1 and level-2 triggers into *full* events. This part of the local journey is at the same time a selection process, because both events and event fragments that do not satisfy the selection criteria are filtered out and discarded from further consideration. The second part of the local data journey, which is carried out by the level-2 trigger, is solely a selection process that filters out the events constructed in the first part that do not satisfy the selection criteria. The third level of the local journey is also solely a selection process that further refines event selections made in previous levels. The above considerations show that during the local journey, events are mutable in the sense that their contents—namely, their constituent signatures—are transformed into full events by the construction and selection processes according to the selection criteria in the trigger menu. Therefore, in the context of the ATLAS experiment, data mutability in the sense of changeabil-

²⁰For more information, see the URL: <https://home.cern/about/computing/grid-system-tiers>

ity of event content is a consequence of data mobility, which is in turn a necessary condition to apply selection criteria and thereby ensure data usability.

The above discussion indicates that the trigger menu used in the ATLAS data acquisition process should also be regarded as the set of event construction criteria, as it serves to construct events out of event fragments. The determination of the trigger menu is partly based on the theoretical judgment that the selection criteria considered relevant to the testing of the predictions of the SM and BSM models should consist of only types of signatures predicted by these models. The determination of the trigger menu also requires a judgment in the form of a data selection strategy, namely that the trigger menu should be sufficiently diversified in terms of types of signatures that are relevant to the intended objectives of the ATLAS experiment. Since the ATLAS experiment also aims at discovering unforeseen phenomena that are not accounted for by the SM and BSM models, the foregoing selection strategy also requires the trigger menu to include selection criteria that are not necessarily based on the predictions of these models. This enables using the same trigger menu to acquire data sets relevant to the entirety of the intended objectives of the ATLAS experiment. The judgment on which the data selection strategy is based is experimental, as it does not follow from the predictions of the SM and BSM models that not dictate how the trigger menu should be diversified in terms of signatures. Therefore, the foregoing theoretical and experimental judgments jointly contribute to the determination of the trigger menu and thereby impose requirements on what counts as usable data in the ATLAS experiment.

The implementation of the above-mentioned experimental strategy in the ATLAS experiment requires taking account of the technical limitations at CERN in terms of data storage capacity and data process time. This in turn leads to the judgment that the trigger menu should be applied to collision events in real time, i.e. while proton collisions are taking place inside the ATLAS detector. This is a technical judgment based on the consideration that the amount of events produced by the LHC is so large that the foregoing technical limitations make it impracticable to apply the trigger menu after events are recorded. It is also experimental in the sense that unlike the experimental judgment concerning the trigger menu, it dictates which specific experimental procedures to use to apply the trigger menu to collision events. It thereby imposes certain technical requirements on the design of the ATLAS data acquisition system. The main technical requirement is the three-level arrangement of the trigger systems in the way it is described in the previous section. There are also more specific requirements concerning the details of the event construction and selection processes. An important technical detail is the use of the seeding mechanism according to which event fragments are accepted to the level-2 trigger in small amounts. If event fragments were accepted at once, this would considerably diminish the level-2 trigger decision time and thus render the level-2 selection process ineffective. The factors such as data processing capacity of each trigger and the amount of events produced by the LHC are also considered in specifying the details of the ATLAS data acquisition system. These technical requirements, together with the ones imposed by the experimental judgments, can be seen as the requirements imposed on the mobility and mutability of the LHC data during its local journey. While the requirements on mobility specify the ways in which events are made to

travel during the construction and selection processes, the requirements on mutability specify the ways in which the contents of events transform during these processes.

In the philosophical literature, the necessity of data mobility and data mutability for data usability has been studied and stressed in relation to data dissemination (see, e.g. Morgan 2010; Leonelli 2015). The present case-study shows that data usability is an essential concern in present-day HEP experiments already in the stage of data acquisition. In this context, in order for the experiment to achieve its intended objectives, it is necessary that the issue of data usability be dealt with before data are disseminated for analysis and interpretation. As the case of the ATLAS experiment illustrates, data mobility and data mutability are necessary conditions to deal with the issue of data usability encountered in data acquisition stage. Thus, in present-day HEP experiments, data does not come ready-made from the detector but rather is constructed to be usable for the purposes of the experiment. As a result of this construction process, data is both mobile and mutable from the outset and prior to its dissemination. Therefore, usability, mobility and mutability are related, though distinct, aspects of data in the context of present-day HEP experiments. What makes these aspects of data related to each other is the fact that they are subjected to some common requirements imposed by theoretical, experimental and technical judgments involved in the design of data acquisition systems.

Acknowledgments I would like to thank the editors of this volume and two anonymous reviewers for their valuable comments and suggestions on earlier drafts of this chapter.

References

- ATLAS Collaboration. 2003. *Technical Design Report: ATLAS High-Level Trigger, Data-Acquisition and Controls*. CERN-LHCC-2003-022.
- . 2012. Observation of a New Particle in the Search for the Standard Model Higgs Boson with the ATLAS Detector at the LHC. *Physics Letters B* 716: 1–29.
- . 2016. Measurement of D^{*+} , D^+ and D_s^{*+} Meson Production Cross Sections in pp Collisions at $\sqrt{s} = 7$ TeV with the ATLAS Detector. *Nuclear Physics B* 907: 717–740.
- Beauchemin, Pierre-Hugues. 2018. Autopsy of Measurements with the ATLAS Detector at the LHC. *Synthese* 194: 275–312.
- Boisot, Max, Markus Nordberg, Saïd Yami, and Bertrand Nicquevert. 2011. *Collisions and Collaboration: The Organization of Learning in the ATLAS Experiment at the LHC*. Oxford: Oxford University Press.
- Branco, Miguel, et al. 2008. Managing ATLAS Data on a Petabyte-Scale with DQ2. *Journal of Physics Conference Series* 119: 062017.
- Boumans, Marcel, and Sabina Leonelli. this volume. From Dirty Data to Tidy Facts: Clustering Practices in Plant Phenomics and Business Cycle Analysis. In *Data Journeys in the Sciences*, ed. Sabina Leonelli and Niccolò Tempini. Cham: Springer.
- CMS Collaboration. 2002. *CMS The TriDAS project: Technical Design Report, Volume 2: Data Acquisition and High-Level Trigger*. CERN-LHCC-2002-026.

- Ellis, Nick. 2010. *Trigger and Data Acquisition*. Lecture given at the 5th CERN-Latin-American School of High-Energy Physics, Recinto Quirama, Colombia, 15–28 Mar 2009. CERN Yellow Report CERN-2010-001, 1–32. <http://lanl.arxiv.org/abs/1010.2942>
- Ellis, John. 2012. Outstanding Questions: Physics Beyond the Standard Model. *Philosophical Transactions of the Royal Society A* 370: 818–830.
- Franklin, Allan. 2013. *Shifting Standards: Experiments in Particle Physics in the Twentieth Century*. Pittsburgh: University of Pittsburgh Press.
- Galison, Peter. 1997. *Image and Logic*. Chicago: University of Chicago Press.
- . 2003. The Collective Author. In *Scientific Authorship: Credit and Intellectual Property in Science*, ed. Peter Galison and Mario Biagioli, 325–353. New York/Oxford: Routledge.
- Gutsche, Oliver, et al. 2017. Big data in HEP: A Comprehensive Use Case Study. *Journal of Physics: Conference Series* 898: 072012.
- Halfmann, Gregor. this volume. Material Origins of a Data Journey in Ocean Science: How Sampling and Scaffolding Shape Data Practices. In *Data Journeys in the Sciences*, ed. Sabina Leonelli and Niccolò Tempini. Cham: Springer.
- Hoeppe, Götz. this volume. Sharing Data, Repairing Practices: On the Reflexivity of Astronomical Data Journeys. In *Data Journeys in the Sciences*, ed. Sabina Leonelli and Niccolò Tempini. Cham: Springer.
- Karaca, Koray. 2017. A Case Study in Experimental Exploration: Exploratory Data Selection at the Large Hadron Collider. *Synthese* 194: 333–354.
- . 2018. Lessons from the Large Hadron Collider for Model-Based Experimentation: The Concept of a Model of Data Acquisition and the Scope of the Hierarchy of Models. *Synthese* 195: 5431–5452.
- Knorr-Cetina, Karin. 1999. *Epistemic Cultures: How the Sciences Make Knowledge*. Cambridge, MA: Harvard University Press.
- Leonelli, Sabina. 2015. What Counts as Scientific Data? A Relational Framework. *Philosophy of Science* 82: 810–821.
- . 2016. *Data-Centric Biology: A Philosophical study*. Chicago: University of Chicago Press.
- . 2019. What Distinguishes Data from Models? *European Journal for Philosophy of Science* 9: 22.
- Morgan, Mary S. 2010. Travelling Facts. In *How Well Do Facts Travel? The Dissemination of Reliable Knowledge*, ed. Peter Howlett and Mary S. Morgan, 3–42. Cambridge: Cambridge University Press.
- Morrison, Margaret. 2015. *Reconstructing Reality: Models, Mathematics, and Simulations*. Oxford: Oxford University Press.

Koray Karaca is Assistant Professor in the Department of Philosophy at the University of Twente in the Netherlands. He received his PhD (2010) in History and Philosophy of Science from Indiana University, Bloomington, USA, and his PhD (2005) in Theoretical Physics from the Middle East Technical University, Ankara, Turkey. Before coming to the University of Twente in September 2015, he worked as a Postdoctoral Researcher between 2010 and 2015 in the interdisciplinary research collaboration “The Epistemology of the Large Hadron Collider”, which is funded by the German Science Foundation and based at the University of Wuppertal, Germany. In the academic year of 2009–2010, he was Visiting Assistant Professor in the Department of Philosophy at the University of South Florida. His research interests include the philosophy and history of modern physics, philosophy of scientific experimentation in the context of high-energy physics experiments and philosophy of modelling, computer simulation and machine learning. His publications appeared in journals such as *The British Journal for the Philosophy of Science*, *Synthese*, *Studies in History and Philosophy of Modern Physics*, *Science in Context* and *Perspectives on Science*.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

