

# Querying Inconsistent Prioritized Data

Camille Bourgaux<sup>1</sup>

<sup>1</sup>DI ENS, ENS, CNRS, PSL University & Inria, 45 rue d'Ulm, 75005 Paris, France

## Abstract

This extended abstract accompanies an invited talk at DL 2024 and presents a summary of some recent results on querying inconsistent prioritized data.

## Keywords


inconsistency handling, prioritized data, optimal repairs

Real-world datasets are plagued by data quality issues which may render the data inconsistent w.r.t. a set of constraints, be they given by database dependencies or ontologies. A prominent way to handle such inconsistent data is to use inconsistency-tolerant semantics to obtain meaningful answers to queries [1, 2, 3]. Most of these semantics are based on the notion of a (subset) repair, which is an inclusion-maximal subset of the data that is consistent with the constraints. In many scenarios, this basic notion of repair can be refined by exploiting preference information about facts. Preferred repairs can then be used in place of subset repairs in any repair-based semantics. This paper presents a summary of some recent results [4, 5, 6] on an approach where preferences are given by a binary priority relation between conflicting facts [7].

**Formal definitions** A knowledge base (KB)  $\mathcal{K} = (\mathcal{D}, \mathcal{T})$  consists of a dataset  $\mathcal{D}$  and a logical theory  $\mathcal{T}$ :  $\mathcal{D}$  is a finite set of facts, and  $\mathcal{T}$  a finite set of first-order logic (FOL) sentences. Typically,  $\mathcal{T}$  is an ontology, e.g., in description logic, or a set of database constraints, e.g., denial constraints of the form  $\forall \vec{x} \neg(\alpha_1 \wedge \dots \wedge \alpha_n)$ , where each  $\alpha_i$  is a relational or inequality atom. A KB  $\mathcal{K} = (\mathcal{D}, \mathcal{T})$  is *consistent* if  $\mathcal{D} \cup \mathcal{T}$  has a model, and *inconsistent* otherwise ( $\mathcal{K} \models \perp$ ). A set of facts  $\mathcal{B}$  is  $\mathcal{T}$ -consistent if  $(\mathcal{B}, \mathcal{T})$  is consistent. A *conflict* of  $\mathcal{K}$  is an inclusion-minimal subset  $\mathcal{C} \subseteq \mathcal{D}$  such that  $(\mathcal{C}, \mathcal{T}) \models \perp$ ;  $\text{Conf}(\mathcal{K})$  denotes the set of conflicts of  $\mathcal{K}$ . A (subset) *repair* of  $\mathcal{K} = (\mathcal{D}, \mathcal{T})$  is an inclusion-maximal  $\mathcal{T}$ -consistent subset  $\mathcal{R} \subseteq \mathcal{D}$ ;  $\text{SRep}(\mathcal{K})$  denotes the set of repairs of  $\mathcal{K}$ . Following [7], we assume that preferences between conflicting facts are available:


**Definition 1.** A priority relation  $\succ$  for a KB  $\mathcal{K} = (\mathcal{D}, \mathcal{T})$  is an acyclic binary relation over  $\mathcal{D}$  such that  $\alpha \succ \beta$  implies  $\{\alpha, \beta\} \subseteq \mathcal{C}$  for some  $\mathcal{C} \in \text{Conf}(\mathcal{K})$ . It is total if for all  $\alpha \neq \beta$  such that  $\{\alpha, \beta\} \subseteq \mathcal{C}$  for some  $\mathcal{C} \in \text{Conf}(\mathcal{K})$ , either  $\alpha \succ \beta$  or  $\beta \succ \alpha$ . A completion of  $\succ$  is a total priority relation  $\succ' \supseteq \succ$ . A prioritized KB  $\mathcal{K}_\succ$  is a KB  $\mathcal{K} = (\mathcal{D}, \mathcal{T})$  with a priority relation  $\succ$  for  $\mathcal{K}$ .


**Definition 2** (Optimal repairs). Let  $\mathcal{K}_\succ$  be a prioritized KB with  $\mathcal{K} = (\mathcal{D}, \mathcal{T})$  and  $\mathcal{R} \in \text{SRep}(\mathcal{K})$ . A Pareto improvement of  $\mathcal{R}$  is a  $\mathcal{T}$ -consistent  $\mathcal{B} \subseteq \mathcal{D}$  such that there is  $\beta \in \mathcal{B} \setminus \mathcal{R}$  with  $\beta \succ \alpha$  for every  $\alpha \in \mathcal{R} \setminus \mathcal{B}$ . A global improvement of  $\mathcal{R}$  is a  $\mathcal{T}$ -consistent  $\mathcal{B} \subseteq \mathcal{D}$  such that  $\mathcal{B} \neq \mathcal{R}$  and for every  $\alpha \in \mathcal{R} \setminus \mathcal{B}$ , there exists  $\beta \in \mathcal{B} \setminus \mathcal{R}$  such that  $\beta \succ \alpha$ . The repair  $\mathcal{R}$  is:

 DL 2024: 37th International Workshop on Description Logics, June 18–21, 2024, Bergen, Norway

 camille.bourgaux@ens.fr (C. Bourgaux)

 0000-0002-8806-6682 (C. Bourgaux)

 © 2024 Copyright for this paper by its author. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)

**Table 1**

Data complexity [7, 4, 5]. Upper bounds hold for conjunctive queries and FOL fragments for which consistency checking and BCQ entailment is in P. Lower bounds hold for atomic queries and any fragment that extends DL-Lite<sub>core</sub> or functional dependencies.

	Globally-optimal	Pareto-optimal	Completion-optimal
AR, IAR	$\Pi_2^P$ -complete	coNP-complete	coNP-complete
Brave	$\Sigma_2^P$ -complete	NP-complete	NP-complete

- Pareto-optimal if there is no Pareto improvement of  $\mathcal{R}$ ;
- globally-optimal if there is no global improvement of  $\mathcal{R}$ ;
- completion-optimal if  $\mathcal{R}$  is a globally-optimal repair of  $\mathcal{K}_{\succ'}$ , for some completion  $\succ'$  of  $\succ$ .

We denote by  $PRep(\mathcal{K}_{\succ})$ ,  $GRep(\mathcal{K}_{\succ})$  and  $CRep(\mathcal{K}_{\succ})$  the sets of Pareto-, globally- and completion-optimal repairs. It holds that  $CRep(\mathcal{K}_{\succ}) \subseteq GRep(\mathcal{K}_{\succ}) \subseteq PRep(\mathcal{K}_{\succ}) \subseteq SRep(\mathcal{K}_{\succ})$ .

If  $\succ$  is induced by assigning scores to facts, then Pareto-, globally- and completion-optimal repairs coincide [8, 9], and also coincide with the  $\subseteq_P$ -repairs from [10].

A Boolean *conjunctive query* (BCQ)  $q$  is entailed by a KB  $\mathcal{K}$  ( $\mathcal{K} \models q$ ) if  $q$  holds in every model of  $\mathcal{K}$ . Hence,  $\mathcal{K} \models \perp$  implies that  $\mathcal{K} \models q$  for every  $q$ . Meaningful answers can be obtained from inconsistent KBs by using *inconsistency-tolerant semantics* based on repairs. The AR semantics defines *plausible* query answers, and is used for consistent query answering in databases [11]. The brave semantics considers all *possible* answers, while the IAR semantics identifies the *most reliable* ones. These semantics can be parametrized by the type of repair they consider.

**Definition 3** (Inconsistency-tolerant semantics). Fix  $X \in \{S, P, G, C\}$  and consider a prioritized KB  $\mathcal{K}_{\succ}$  with  $\mathcal{K} = (\mathcal{D}, \mathcal{T})$  and a BCQ  $q$ . Then  $q$  is entailed by  $\mathcal{K}_{\succ}$  under

- X-AR semantics, denoted  $\mathcal{K}_{\succ} \models_{AR}^X q$ , if  $(\mathcal{R}, \mathcal{T}) \models q$  for every  $\mathcal{R} \in XRep(\mathcal{K}_{\succ})$ ;
- X-brave semantics, denoted  $\mathcal{K}_{\succ} \models_{brave}^X q$ , if  $(\mathcal{R}, \mathcal{T}) \models q$  for some  $\mathcal{R} \in XRep(\mathcal{K}_{\succ})$ ;
- X-IAR semantics, denoted  $\mathcal{K}_{\succ} \models_{IAR}^X q$ , if  $(\mathcal{B}, \mathcal{T}) \models q$  where  $\mathcal{B} = \bigcap_{\mathcal{R} \in XRep(\mathcal{K}_{\succ})} \mathcal{R}$ .

The semantics are related as follows:  $\mathcal{K}_{\succ} \models_{IAR}^X q \Rightarrow \mathcal{K}_{\succ} \models_{AR}^X q \Rightarrow \mathcal{K}_{\succ} \models_{brave}^X q$ .

Table 1 gives the data complexity of BCQ entailment under optimal repair-based semantics.<sup>1</sup> Note that S-AR entailment is coNP-hard even in very basic settings [12, 13], while S-brave or S-IAR entailment is tractable for DL-Lite ontologies or denial constraints [14].

**Practical SAT-based approaches** [5] The (co)NP complexity results naturally suggest the interest of employing *SAT solvers* to compute query answers under X-AR, X-brave and X-IAR semantics for  $X \in \{P, C\}$ , as it has already been done for S-AR [15, 16]. There are several ways of employing SAT encodings to compute answers under (optimal) repair-based semantics, by exploiting different reasoning modes of SAT solvers. The ORBITS system implements different algorithms and encodings for the case where conflicts are binary. It takes as input the conflicts,

<sup>1</sup>The lower bound for G-IAR and G-brave with functional dependencies follows from the proof of [7, Theorem 2].

priority relation, and the potential query answers associated with their *causes*, where a cause for  $q$  is a minimal  $\mathcal{T}$ -consistent subset  $\mathcal{C} \subseteq \mathcal{D}$  such that  $(\mathcal{C}, \mathcal{T}) \models q$ . Our experimental comparison shows that the choice of algorithm and encoding variant may have huge impact on the computation time. Moreover, while in some cases our results can be used to single out some approaches as more effective, more often there are no clear winner(s). For example, we consider two ways of encoding the absence of a cause, and choosing one instead of the other may make a notable difference, but which one is best depends on the KB and query.

**Connections with abstract argumentation [4]** An *argumentation framework (AF)* is a pair  $(Args, \rightsquigarrow)$  where  $Args$  is a finite set of arguments and  $\rightsquigarrow \subseteq Args \times Args$  is the attack relation:  $\alpha$  attacks  $\beta$  if  $\alpha \rightsquigarrow \beta$ . The semantics of AFs is based on *extensions*, which can be seen as coherent sets of arguments. Many notions of extension have been considered, in particular *preferred extensions* and *stable extensions* [17]. When these two kinds of extension coincide, the AF is called *coherent*. We define *preference-based set-based argumentation framework (PSETAF)*, a variant of AF based on collective attacks with a preference relation between arguments. We exhibit a natural translation of a prioritized KB  $\mathcal{K}_{\succ}$  into a PSETAF  $F_{\mathcal{K}, \succ}$  that uses  $\mathcal{D}$  as the arguments and show that  $\mathcal{R} \in PRep(\mathcal{K}_{\succ})$  iff it is a stable extension of  $F_{\mathcal{K}, \succ}$ . Moreover, if  $\succ$  is transitive or if  $\mathcal{K}$  has only binary conflicts, then  $F_{\mathcal{K}, \succ}$  is coherent so  $\mathcal{R} \in PRep(\mathcal{K}_{\succ})$  iff it is a preferred extension of  $F_{\mathcal{K}, \succ}$ . Since there is no notion of extension that corresponds to globally- or completion-optimal repairs, this speaks to the interest of adopting Pareto-optimal repairs.

The argumentation connection allows us to propose a new notion of grounded repair, directly inspired by the grounded extension from argumentation. The (unique) grounded repair is contained in the intersection of Pareto-optimal repairs and can be computed in polynomial time from the dataset and conflicts. Moreover, it is more productive than the Elect semantics [18] and lies between the non-defeated and the prioritized inclusion-based non-defeated repairs that have been proposed in the case where the priority relation is score structured [19].

**Connections with active integrity constraints [6]** In the database setting, *active integrity constraints (AICs)* state how to resolve constraint violations [20, 21]. A ground AIC is a formula of the form  $\alpha_1 \wedge \dots \wedge \alpha_n \wedge \neg\beta_1 \wedge \dots \wedge \neg\beta_m \rightarrow \{A_1, \dots, A_k\}$  where the  $\alpha_j, \beta_j$  are facts and each  $A_i$  is an *update action* of the form  $-\alpha_j$  for some  $1 \leq j \leq n$  or  $+\beta_j$  for some  $1 \leq j \leq m$ . The semantics of a set of AICs is based on the notion of a *repair update*, defined as a minimal set  $\mathcal{U}$  of update actions such that  $\mathcal{D} \setminus \{\alpha \mid -\alpha \in \mathcal{U}\} \cup \{\alpha \mid +\alpha \in \mathcal{U}\}$  satisfies the constraints. Many proposals have been made to select the appropriate repair updates to take into account a set of AICs, such as *founded*, *well-founded*, *grounded* and *justified* repair updates.

Given a prioritized database (i.e., a prioritized KB such that  $\mathcal{T}$  is a set of database constraints), we construct a set of ground AICs such that the Pareto-optimal repairs coincide with the repairs obtained by applying founded, grounded and justified repair updates w.r.t. the generated set of AICs. We also exhibit a translation of a ‘well-behaved’ set of AICs into a prioritized database and again relate Pareto-optimal repairs with founded, grounded and justified repair updates. We take this as further evidence that Pareto-optimal repairs are an especially natural notion.

These results hold not only for denial constraints but also for *universal constraints* of the form  $\forall \vec{x} \neg(\alpha_1 \wedge \dots \wedge \alpha_n \wedge \neg\beta_1 \wedge \dots \wedge \neg\beta_m)$ , where repairs can also be obtained by *adding facts* (while minimizing the symmetric difference with the original database). Such constraints and repairs could be explored in the KB setting, e.g., for ontologies with closed predicates [22].

## Acknowledgments

I thank the DL 2024 workshop chairs for inviting me. The work summarized in this extended abstract has been conducted together with Meghyn Bienvenu (CNRS & University of Bordeaux) and supported by the ANR AI Chair INTENDED (ANR-19-CHIA-0014).

## References

- [1] M. Bienvenu, C. Bourgaux, Inconsistency-tolerant querying of description logic knowledge bases, in: Tutorial Lectures of the 12th International Reasoning Web Summer School, 2016.
- [2] M. Bienvenu, A short survey on inconsistency handling in ontology-mediated query answering, *Künstliche Intelligenz* 34 (2020) 443–451.
- [3] J. Wijsen, Foundations of query answering on inconsistent databases, *SIGMOD Record* 48 (2019) 6–16.
- [4] M. Bienvenu, C. Bourgaux, Querying and repairing inconsistent prioritized knowledge bases: Complexity analysis and links with abstract argumentation, in: Proceedings of the 17th International Conference on Principles of Knowledge Representation and Reasoning (KR), 2020.
- [5] M. Bienvenu, C. Bourgaux, Querying inconsistent prioritized data with ORBITS: algorithms, implementation, and experiments, in: Proceedings of the 19th International Conference on Principles of Knowledge Representation and Reasoning (KR), 2022.
- [6] M. Bienvenu, C. Bourgaux, Inconsistency handling in prioritized databases with universal constraints: Complexity analysis and links with active integrity constraints, in: Proceedings of the 20th International Conference on Principles of Knowledge Representation and Reasoning, (KR), 2023.
- [7] S. Staworko, J. Chomicki, J. Marcinkowski, Prioritized repairing and consistent query answering in relational databases, *Annals of Mathematics and Artificial Intelligence (AMAI)* 64 (2012) 209–246.
- [8] E. Livshits, B. Kimelfeld, Counting and enumerating (preferred) database repairs, in: Proceedings of the 36th ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems (PODS), 2017.
- [9] C. Bourgaux, Inconsistency Handling in Ontology-Mediated Query Answering. (Gestion des incohérences pour l'accès aux données en présence d'ontologies), Ph.D. thesis, University of Paris-Saclay, France, 2016.
- [10] M. Bienvenu, C. Bourgaux, F. Goasdoué, Querying inconsistent description logic knowledge bases under preferred repair semantics, in: Proceedings of the 28th AAAI Conference on Artificial Intelligence (AAAI), 2014.
- [11] L. E. Bertossi, Database Repairing and Consistent Query Answering, *Synthesis Lectures on Data Management*, Morgan & Claypool Publishers, 2011.
- [12] S. Arming, R. Pichler, E. Sallinger, Complexity of repair checking and consistent query answering, in: Proceedings of the 19th International Conference on Database Theory (ICDT), 2016.
- [13] M. Bienvenu, On the complexity of consistent query answering in the presence of simple

- ontologies, in: Proceedings of the 26th AAAI Conference on Artificial Intelligence (AAAI), 2012.
- [14] M. Bienvenu, R. Rosati, Tractable approximations of consistent query answering for robust ontology-based data access, in: Proceedings of the 23rd International Joint Conference on Artificial Intelligence (IJCAI), 2013.
  - [15] M. Bienvenu, C. Bourgaux, F. Goasdoué, Computing and explaining query answers over inconsistent DL-Lite knowledge bases, *Journal of Artificial Intelligence Research (JAIR)* 64 (2019) 563–644.
  - [16] A. A. Dixit, P. G. Kolaitis, A SAT-based system for consistent query answering, in: Proceedings of the 22nd International Conference on Theory and Applications of Satisfiability Testing (SAT), 2019.
  - [17] P. M. Dung, On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games, *Artif. Intell.* 77 (1995) 321–358.
  - [18] S. Belabbes, S. Benferhat, J. Chomicki, Elect: An inconsistency handling approach for partially preordered lightweight ontologies, in: Proceedings of the 15th International Conference on Logic Programming and Nonmonotonic Reasoning (LPNMR), 2019.
  - [19] S. Benferhat, Z. Bouraoui, K. Tabia, How to select one preferred assertional-based repair from inconsistent and prioritized DL-Lite knowledge bases?, in: Proceedings of the 24th International Joint Conference on Artificial Intelligence (IJCAI), 2015.
  - [20] S. Flesca, S. Greco, E. Zumpano, Active integrity constraints, in: Proceedings of the 6th International ACM SIGPLAN Conference on Principles and Practice of Declarative Programming (PPDP), 2004.
  - [21] B. Bogaerts, L. Cruz-Filipe, Fixpoint semantics for active integrity constraints, *Artif. Intell.* 255 (2018) 43–70.
  - [22] C. Lutz, I. Seylan, F. Wolter, The data complexity of ontology-mediated queries with closed predicates, *Log. Methods Comput. Sci.* 15 (2019).