# On the Local Closed-World Assumption of Data-Sources [1]

Alvaro Cortés-Calabuig [a] Marc Denecker [a] Ofer Arieli [b]
Bert Van Nuffelen [a] Maurice Bruynooghe [a]

[a] *Department of Computer Science, Katholieke Universiteit Leuven, Belgium.*
[b] *Department of Computer Science, The Academic College of Tel-Aviv, Israel.*

## 1 Introduction

The *Closed-World Assumption* (CWA) on a database expresses that an atom not in the database is false. The CWA is only applicable in domains where the database has complete knowledge. In many cases, for example in the context of distributed databases, a data source has only complete knowledge about part of the domain of discourse. In this paper, we introduce an expressive and intuitively appealing method of representing a *local closed-world assumption* (LCWA) of autonomous data-sources. This approach distinguishes between the data that is conveyed by a data-source and the meta-knowledge about the area in which the data is complete. The data is stored in a relational database that can be queried in the standard way, whereas the meta-knowledge about its completeness is expressed by a first order theory that can be processed by an independent reasoning system (for example a *mediator*).The full version of this work appears in [1], where we consider different ways of representing our approach, relate it to other methods of representing local closed-word assumptions of data-sources, and show some useful properties of our framework which facilitate its application in real-life systems.

*Example* 1. Consider a distributed traffic tax administration system, in which there is one data-source for each county, maintaining a database of car owners in that county. There is a protocol amongst the different counties so that when a car owner leaves one county $\mathcal{A}$ to live in another county $\mathcal{B}$, then county $\mathcal{A}$ immediately transfers its information to county $\mathcal{B}$, while still preserving a record of the car owner and its current status for a certain period of time, to handle all running tax demands. By the nature of the protocol, we may assume that each data-source has complete knowledge about all car owners in its county, but in general it has more information than that. Part of the tables of a particular county, say Bronx, may look as follows:

Car Owners

| Name | Model | CarID |
|---|---|---|
| Peter Steward | Mercedes 320 | Qn5452 |
| John Smith | Volvo 230 | Bx5242 |
| Mary Clark | BMW 550 | Bx5462 |

Location

| Name | Residence |
|---|---|
| Peter Steward | Queens |
| Mary Clark | Bronx |
| John Smith | Bronx |

By the nature of the distributed system, this data-source has an expertise on car owners of Bronx. This meta-knowledge allows to derive further information that is not explicitly stated in the data-source, e.g. that all people that are recorded in the table Location as residents of Bronx, are actually *all* the car owners from that county. However, as the information about car owners in Queens is not complete in this data-source, one should not rely only on the tables of this source for making further conclusions about that county.

[1] This work is supported by FWO-Vlaanderen, European Framework 5 Project WASP, and by GOA/2003/08.

## 2 The Local Closed-World Assumption

**Definition 1.** A *data-source* $S$ is a pair $\langle \Sigma, D\rangle$, where $\Sigma$ is a vocabulary consisting of predicate symbols in a fixed relational schema $\mathcal{R}(\Sigma)$ and a finite set $\mathcal{C}(\Sigma)$ of constants representing the elements of the domain of discourse; and $D$ is a finite set of ground atoms expressed in terms of $\Sigma$.

**Definition 2.** A *local closed-world assumption* for a data-source $S = \langle \Sigma, D\rangle$, is a triple $\mathcal{LCWA} = \langle S, \overline{P}, \Psi[\overline{x}]\rangle$, where $\overline{P} = \{P_1(\overline{x}_1), \ldots, P_n(\overline{x}_n)\}$ is a set of atoms (the LCWA's objects) and $\Psi[\overline{x}]$ (the context of the assumption) is a first-order formula over $\Sigma$, $\overline{x} = \bigcup_{i=1}^{n} \overline{x}_i$, and $\Psi[\overline{x}]$ denotes that the free variables of $\Psi$ are a subset of $\overline{x}$.

*Example* 2. Let $S = \langle \Sigma, D\rangle$ the formal representation of the data-source in Example 1. The assumption $\langle S, \{\mathsf{CarO}(x, y, u)\}, \mathsf{Loc}(x, \mathsf{Bx})\rangle$ expresses that $S$ contains all the data about the cars of persons living in $\mathsf{Bx}$.

**Definition 3.** Let $\mathcal{LCWA} = \langle S, \{P_1(\overline{x}_1), \ldots, P_n(\overline{x}_n)\}, \Psi[\overline{x}]\rangle$ be a local closed-world assumption for a data-source $S$. Denote by $P^D$ the set of tuples of $P$ in $D$. $P(\overline{t}) \in P^D$, where $\overline{t}$ is a tuple of terms, abbreviates the formula $\bigvee_{\overline{a} \in P^D} (\overline{t} = \overline{a})$. The formula that is *induced* from $\mathcal{LCWA}$, denoted by $\Lambda_{\mathcal{LCWA}}$, is the following:

$$\forall \overline{x}\Big(\Psi[\overline{x}] \to \bigwedge_{i=1}^{n} \big(P_i(\overline{x}_i) \to \big(P_i(\overline{x}_i) \in P_i^D\big)\big)\Big),$$

where $\overline{x} = \bigcup_{i=1}^{n} \overline{x}_i$.

*Example* 3. Below is the formula that is induced from the local closed-world assumption in Example 2.

$$\begin{aligned}\forall x \forall y \forall u \big(\mathsf{Loc}(x, \mathsf{Bx}) \to (\mathsf{CarO}(x, y, u) \to ((&x = \mathsf{PS} \wedge y = \mathsf{M320} \wedge u = \mathsf{Qn5452}) \vee \\ &(x = \mathsf{JS} \wedge y = \mathsf{V230} \wedge u = \mathsf{Bx5242}) \vee \\ &(x = \mathsf{MC} \wedge y = \mathsf{B550} \wedge u = \mathsf{Bx5462}))\big)\big)\end{aligned}$$

**Definition 4.** For a data-source $S = \langle \Sigma, D\rangle$, denote: $\mathfrak{D}(S) = \bigwedge_{d \in D} d$.

Now we are ready to define the meaning of a data-source (in the context of mediator systems):

**Definition 5.** Let $S = \langle \Sigma, D\rangle$ be a data-source and let $\mathcal{LCWA}^j = \langle S, \overline{P}^j, \Psi^j\rangle$, $j = 1, \ldots, m$, be all the local closed-world assumptions of $S$. Then the *meaning* of $S$ is given by the following formula:

$$\mathfrak{M}(S) = \mathfrak{D}(S) \wedge \bigwedge_{j=1}^{m} \Lambda_{\mathcal{LCWA}^j}.$$

The full version of this work appears in [1].

## References


[1] Proc. 8th international conference on logic programming and non-monotonic reasoning. In C. Baral, G. Greco, N. Leone, and G. Tarracina, editors, *LPNMR'05*, volume 3662 of *LNAI*, pages 145–157. Springer, 2005.