$\mathcal{M}$olecules $\mathcal{o}$f $\mathcal{K}$nowledge

Self-Organisation in Knowledge-Intensive Environments

Stefano Mariani

s.mariani@unibo.it

DISI
Alma Mater Studiorum—Università di Bologna

Seminar @ DISI, Cesena - November 13th, 2014

Outline

Context

Knowledge-Intensive Environments

Knowledge-Intensive Environments (KIE) [Bhatt, 2001]: "systems" combining business processes, technologies and people's skills to store, handle, make accessible – in one word, *manage* – very large repositories of information — e.g. wiki portals, online press, enterprise social networks, etc.

Peculiar challenges from the infrastructural standpoint:

- data size — from GBs to TBs
- scale — from organization-wide to world-wide
- dynamism — new information produced/consumed at fast pace — e.g. tweets
- diversity — both in information representation and usage destination
- openness — new users can enter/leave the system at any time
- unpredictability — KIE are often *socio-technical systems*, thus involving humans, whose behaviour is rarely fully predictable

Motivations

KIE challenges are usually faced using *brute force* approaches relying on ever-increasing (hopefully, endless) (i) computational power and (ii) storage

- "big data" techniques, non-relational large-scale DBs, "data-in-the-cloud" paradigm, other buzzwords ;)

! This won't scale forever !

E.g. what about the end of Moore's law?

"Dead data"

One possible research line departs from the following question: why do we stick to view data as a passive, "dead" thing to run algorithms upon in the traditional I/O paradigm?

Goals

The *M*olecules *of* *K*nowledge approach

Data is alive, information is a living thing *continuously* and *spontaneously* interacting with other information as well as with its *prosumers*, evolving itself according to such interactions [Mariani, 2011].

*M*olecules *of* *K*nowledge (*MoK*) [Mariani and Omicini, 2013] promotes this interpretation by relying on the following features to tackle KIE challenges:

- opportunism — "trial-and-error" rather than brute-force search
- locality — partial information, *decentralized* algorithms, local interactions
- probability — don't strive to predict the unpredictable: rely on probability
- situatedness — don't try to account for every possible scenario: be ready to adapt to the current situation

MoK leads to self-organising knowledge management, making "knowledge structures" appear *by emergence* from local interactions among "live" data chunks [Mariani and Omicini, 2012].

Outline

1. Context, Motivations & Goals
2. Molecules of Knowledge (MoK)
   - Vision
   - Model
3. Evaluating MoK
   - Case Study
4. Conclusion & Open Questions

Envisioning MoK I

Definition

Molecules of Knowledge is a coordination model promoting self-organisation of knowledge in multi-agent systems (MAS), toward the idea of "self-organising workspaces" [Omicini, 2011].

Main goals:

- *autonomously* aggregate data to build more "complex" heaps of information — possibly conveying *novel* knowledge previously unknown or hidden
- *autonomously* spread such information toward potentially interested knowledge prosumers — rather than be searched *proactively*

Main "sources of inspiration":

biochemical coordination — in particular, Biochemical Tuple Spaces [Viroli and Casadei, 2009]

stigmergic coordination — in particular, Behavioural Implicit Communication [Castelfranchi et al., 2010]

Envisioning MoK II

- Why a coordination model to manage data?
  - → if data is alive, then we need to properly *coordinate* such "living" data chunks' interactions [Omicini and Viroli, 2011]
- Why biochemical coordination?
  - → the *chemical metaphor* has been already shown to effectively deal with scale, openness and data size issues in MAS in a simple yet efficient way, by leveraging *locality* and *probability* features [Viroli and Casadei, 2009, Zambonelli et al., 2011]
- Why stigmergic coordination?
  - → the power of *environment-mediated* interactions in MAS has been already shown to successfully deal with diversity, dynamism and unpredictability, by leveraging on *situatedness* [Weyns et al., 2007, Castelfranchi et al., 2010]

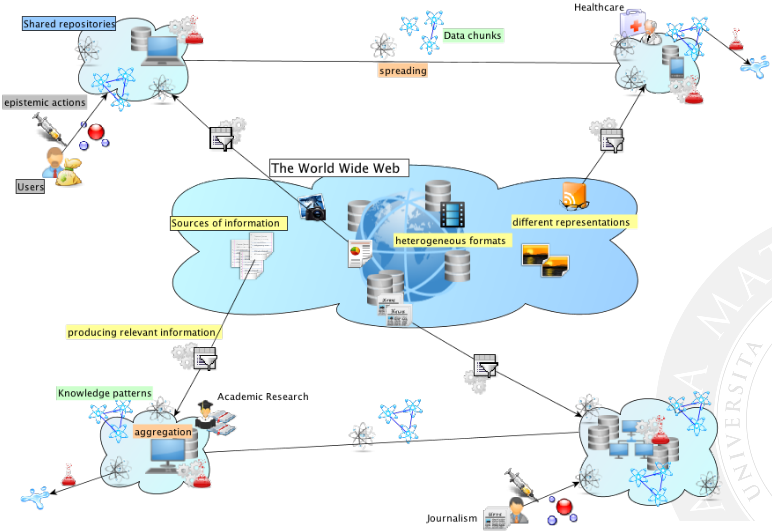
Envisioning MoK III

- A MoK system should be seen as a network of shared information repositories, in which some source entities continuously and spontaneously put data chunks
- Such data may then (i) aggregate so as to reify some (potentially) relevant "knowledge-related patterns"
  - e.g. linking two news stories talking about the same person or written by the same author, read by the same prosumer or both related to a third news story

  (ii) diffuse among these networked shared spaces toward the (potentially) interested users
  - e.g. papers about MAS should strive to reach MAS researchers' repositories
- Users can interact with the system through epistemic actions
  - e.g. read a post, contribute to a wiki, highlight words in an article, ...

  which are tracked and exploited by the MoK system to influence knowledge evolution transparently to the user
  - e.g., a user highlighting a given word may imply such user being highly interested in such topics, thus MoK can react by, e.g., increasing rank position of related topics in a search query

A *MoK* System I



"Contact Point"

MoK as Pervasive Information Fields ?

- *MoK* (networked) compartments can be interpreted as pervasive information fields, diffusing information from neighbourhood to neighbourhood. . . ?
- *MoK* (networked) compartments can be interpreted as the agents' environment, affected and influencing agents' behaviour based on such information fields. . . ?

Overview

The Molecules of Knowledge model features the following abstractions [Mariani and Omicini, 2013]:

- **Seeds** — The "sources" of information
- **Compartments** — The information repositories all other MoK abstractions (conceptually) belong to
- **Catalysts** — "Prosumers" of information (both producer and consumer users)
- **Atoms** — The "primitive" unit of information in MoK (always produced by a seed)
- **Molecules** — The "composite" unit of information in MoK (interaction "patterns" found by MoK itself among different atoms are reified into molecules)
- **Enzymes** — The reification of catalysts' epistemic actions
- **Reactions** — The "laws of nature" driving MoK compartments evolution

Definitions I

Seeds

$\mathcal{MoK}$ Seeds are the sources of information, *continuously* and *spontaneously* injecting data pieces (atoms) into the workspace (compartment) they belong to.

Compartments

$\mathcal{MoK}$ Compartments are the *active* repositories of information storing seeds, atoms, molecules and enzymes, also responsible for scheduling and executing $\mathcal{MoK}$ reactions.

Catalysts

$\mathcal{MoK}$ Catalysts are the users of the $\mathcal{MoK}$ system, the *prosumers* both exploiting and influencing $\mathcal{MoK}$ self-organisation services.

Definitions II

Atoms

Atoms are the smallest, "atomic" unit of information in $\mathcal{M}o\mathcal{K}$. They are *continuously* and *spontaneously* injected by a seed into the compartment it belongs to.

Molecules

Molecules are the evolving, composite unit of information in $\mathcal{M}o\mathcal{K}$. They are *continuously* and *spontaneously* produced by the compartment they belong to so as to reify some meaningful knowledge.

Enzymes

Enzymes are the reification of catalysts' epistemic actions over the information stored within their compartment. They are *automatically* produced by the compartment to be exploited into $\mathcal{M}o\mathcal{K}$ reactions.

Definitions III

Reactions

Reactions are the *general purpose coordination laws* resembling biochemical reactions in $\mathcal{MoK}$. They drive information evolution within the compartment they are "installed" in by governing molecules interactions.

Four reactions are considered so far[1] in the $\mathcal{MoK}$ model:

- **Aggregation** — Bounds together semantically related molecules
- **Diffusion** — Moves molecules between neighbour compartments
- **Reinforcement** — Consumes enzymes to increase the concentration of their molecule
- **Decay** — Decreases concentration of molecules as time passes

[1] This is by no means the definite set of available reactions. Work on studying the expressiveness of this "core" reactions set is in progress — see FoCAS2014 paper, to appear online soon as SASO workshop proceedings.

A $\mathcal{MoK}$ System II

- A $\mathcal{MoK}$ system is a network of $\mathcal{MoK}$ compartments, in which $\mathcal{MoK}$ seeds continuously and spontaneously inject $\mathcal{MoK}$ atoms
- $\mathcal{MoK}$ atoms (and eventually $\mathcal{MoK}$ molecules) may then aggregate, diffuse, reinforce and decay, driven by $\mathcal{MoK}$ reactions and $\mathcal{MoK}$ enzymes produced by $\mathcal{MoK}$ catalysts' behaviour
- $\mathcal{MoK}$ reactions are scheduled and executed by $\mathcal{MoK}$ compartments according to Gillespie's *chemical dynamics* simulation algorithm [Gillespie, 1977], so as to promote self-organisation based on locality, situatedness and probability

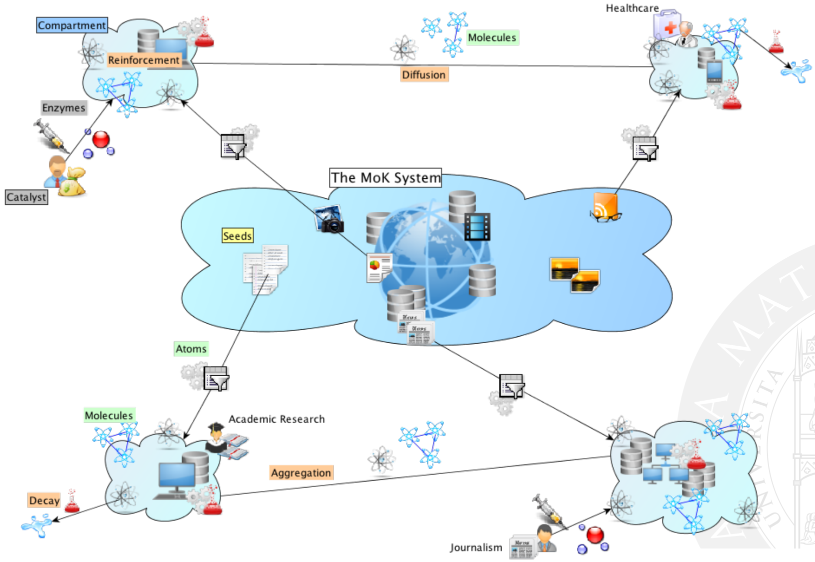
  probability — matching of reactants in $\mathcal{MoK}$ reactions is probabilistically based on the $\mathcal{F}_{\mathcal{MoK}}$ function[2]; scheduling of reactions is probabilistic according to Gillespie algorithm

  situatedness — decay and diffusion enforce, respectively, situatedness in time and space; reinforcement supports situatedness "in context"

  locality — compartments execute $\mathcal{MoK}$ reactions locally; diffusion is based on neighbourhood

[2] $\mathcal{MoK}$ semantic similarity measure, representing the extent to which different atoms/molecules are semantically related.

A MoK System III



Outline

$\mathcal{M}o\mathcal{K}$ Early Stage Evaluation

- To date, $\mathcal{M}o\mathcal{K}$ is still little more than an (hopefully, interesting) idea: no large-scale $\mathcal{M}o\mathcal{K}$ systems have been deployed in the real world, no large-scale simulations of all $\mathcal{M}o\mathcal{K}$ features have been carried out
- Nevertheless, a prototype implementation of a "$\mathcal{M}o\mathcal{K}$ engine" exists, deployed upon the TuCSoN coordination infrastructure[3] [Omicini and Zambonelli, 1999]
  - Roughly speaking, TuCSoN tuple centres [Omicini and Denti, 2001] reify $\mathcal{M}o\mathcal{K}$ compartments and a ReSpecT specification [Omicini, 2007] therein installed implements $\mathcal{M}o\mathcal{K}$ reactions
- Also, simulations regarding $\mathcal{M}o\mathcal{K}$ reactions interplay and effectiveness in exhibiting self-organising behaviours have been carried out
  - In particular, using the Bio-PEPA framework [Ciocchetta and Hillston, 2009] to face the problem of *parameter engineering* in $\mathcal{M}o\mathcal{K}$ reactions' rate expressions [Mariani, 2013]

[3] Home page at `http://tucson.unibo.it`

The $\mathcal{MoK}$-News Scenario I

Why News?

News management systems are a prominent example of socio-technical, knowledge intensive environments.

heterogeneity — news "sources" can be virtually *anything* — press, social networks, enterprise reports, ...

ubiquity — netbooks, tablets and smartphones made information production, sharing and consumption as *pervasive* as never before

unpredictability — while journalists can follow some *structured* workflow, the same does not hold for the "citizen journalism"

$\mathcal{MoK}$ & News

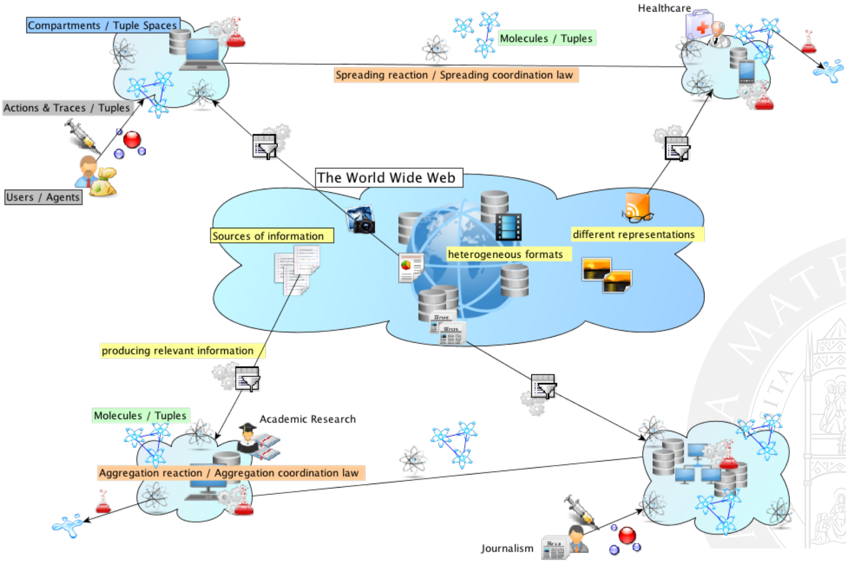
New management systems are thus well suited for $\mathcal{MoK}$.

The MoK–News Scenario II

- In [Mariani and Omicini, 2012], we took IPTC's[4] technical standards regarding news management, in particular:
  - `NewsML` — an *XML-based* tagging language meant to ease news sharing by relying on `NewsCodes` *ontologies*
  - `NITF` — an XML-based tagging language meant to enrich the news content, again by relying on `NewsCodes`
- We identified an abstract mapping between a MoK atom and a `NewsML`/`NITF` *tag*, in particular:
  - → `atom`(*src*, *val*, `sem`(*tag*, *catalog*))$_c$ where *tag* ::= `NewsML`/`NITF` *tag* and *catalog* ::= `NewsCodes` *uri*
- We implemented the domain-specific MoK model – called MoK–`News` – on an existing platform for distributed coordination in MAS, in particular:
  - → on TuCSoN, by using its tuple centres as MoK compartments, in which a ReSpecT specification implements the Gillespie algorithm; TuCSoN tuples as MoK seeds/atoms/molecules/enzymes/reactions and TuCSoN agents as MoK catalysts

[4] `http://www.iptc.org/site/Home/About/`

A $\mathcal{M}o\mathcal{K}$ System IV



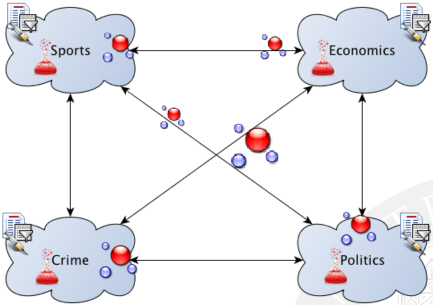
A $\mathcal{M}o\mathcal{K}$ System V

A $\mathcal{M}o\mathcal{K}$-`News` system could then be deployed to an online magazine newsroom as follows:

- journalists may be given a smartphone to use as their workspace, running a $\mathcal{M}o\mathcal{K}$ compartment decorated with a suitable GUI letting them carry out their usual working habits — searching for news in the web, storing some for later use, creating and editing stories, ...
- within compartments, news sources (seeds), news pieces (atoms/molecules) and journalists' actions traces (enzymes) live and interact in a completely autonomous way so as to best support journalists' workflow — seeds increasing/decreasing atoms injection rate, molecules aggregating to automagically compose news stories, enzymes increasing/decreasing relevance (concentration) of news pieces, ...
- as they use the system, journalists implicitly drive its behaviour (e.g. reaction rates) toward their needs, *closing the feedback loop* providing $\mathcal{M}o\mathcal{K}$-`News` with self-adaptive capabilities — combination of diffusion and reinforcement, driven by journalists' enzymes, enables smart migration

"Smart Migration" in $\mathcal{MoK}$-`News` I

- Imagine the $\mathcal{MoK}$-`News` compartments topology to the right to be deployed
- At "bootstrap", news stories are equally distributed among the compartments — since no *a priori* knowledge about journalists' interest topics is assumed
- diffusion and decay reactions *apply to all* molecules with *same rate expression*



! Eventually, the news "spatial organisation" will change according to journalists' interactions — in particular, thanks to the interplay between enzymes and $\mathcal{MoK}$ diffusion, reinforcement and decay reactions

"Smart Migration" in $\mathcal{M}o\mathcal{K}$-News II

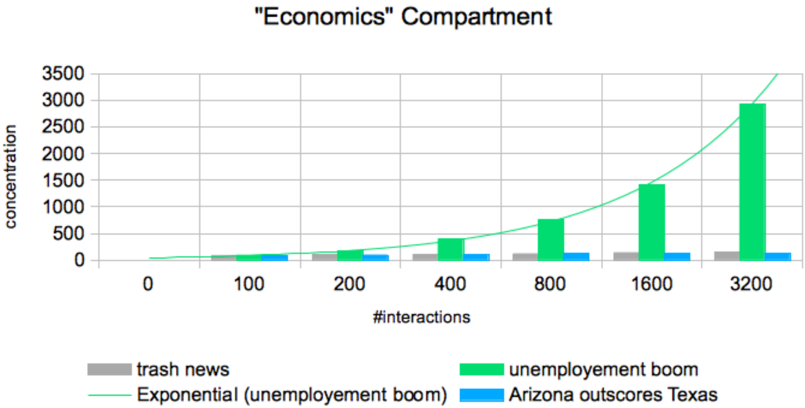


Figure : "Economics news" tend to be the most numerous (hence, relevant) in the "economics compartment" — that is, the one belonging to a journalist interested in economics.

"Smart Migration" in $\mathcal{M}o\mathcal{K}$-News III

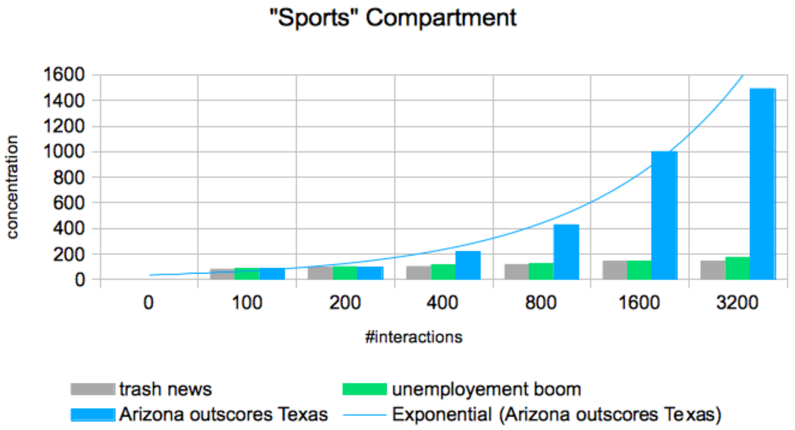


Figure : In another compartment (e.g. "sports compartment"), they are not relevant at all, because the journalist working therein is focussed on other topics (e.g. "sports news")

Outline

Conclusion

- The $\mathcal{MoK}$ model for self-organisation of knowledge promotes a novel interpretation of information as "living things" autonomously interacting, moving, evolving
- Early simulations have confirmed some of $\mathcal{MoK}$'s desiderata regarding run-time behaviour of its reactions
- Early evaluation on a case study implemented on top of a prototype "$\mathcal{MoK}$ engine" also confirmed such desiderata, in particular, exhibiting a self-organising behaviour regarding spatial displacement of information

Open Questions

- Implementing a full-fledged $\mathcal{MoK}$ infrastructure is still far from current $\mathcal{MoK}$ state — despite the prototype
- Many skills are required to "do the job" — e.g. knowledge representation and storage, semantic matching, data mining, self-organisation mechanisms, middleware, ...
- Also the $\mathcal{MoK}$ model needs to be completed and more formally investigated, especially about its expressiveness in reaching self-* behaviours — recent work in this direction, see FoCAS2014 paper.

References I

Bhatt, G. D. (2001).
Knowledge management in organizations: examining the interaction between technologies, techniques, and people.
Journal of Knowledge Management, 5(1):68–75.

Castelfranchi, C., Pezzullo, G., and Tummolini, L. (2010).
Behavioral implicit communication (BIC): Communicating with smart environments via our practical behavior and its traces.
International Journal of Ambient Computing and Intelligence, 2(1):1–12.

Ciocchetta, F. and Hillston, J. (2009).
Bio-PEPA: A framework for the modelling and analysis of biological systems.
Theoretical Computer Science, 410(33–34):3065 – 3084.
Concurrent Systems Biology: To Nadia Busi (1968–2007).

Gillespie, D. T. (1977).
Exact stochastic simulation of coupled chemical reactions.
The Journal of Physical Chemistry, 81(25):2340–2361.

References II

Mariani, S. (2011).
Molecules of knowledge: a new approach to knowledge production, management and consumption.
Master's thesis, II Facoltà di Ingegneria Informatica – Alma Mater Studiorum Università di Bologna.

Mariani, S. (2013).
Parameter engineering vs. parameter tuning: the case of biochemical coordination in MoK.
In Baldoni, M., Baroglio, C., Bergenti, F., and Garro, A., editors, *From Objects to Agents*, volume 1099 of *CEUR Workshop Proceedings*, pages 16–23, Turin, Italy. Sun SITE Central Europe, RWTH Aachen University.
XIV Workshop (WOA 2013). Workshop Notes.

Mariani, S. and Omicini, A. (2012).
Self-organising news management: The *Molecules of Knowledge* approach.
In Fernandez-Marquez, J. L., Montagna, S., Omicini, A., and Zambonelli, F., editors, *1st International Workshop on Adaptive Service Ecosystems: Natural and Socially Inspired Solutions (ASENSIS 2012)*, pages 11–16, SASO 2012, Lyon, France.
Pre-proceedings.

References III

Mariani, S. and Omicini, A. (2013).
Molecules of Knowledge: Self-organisation in knowledge-intensive environments.
In Fortino, G., Bădică, C., Malgeri, M., and Unland, R., editors, *Intelligent Distributed Computing VI*, volume 446 of *Studies in Computational Intelligence*, pages 17–22. Springer.
6th International Symposium on Intelligent Distributed Computing (IDC 2012), Calabria, Italy, 24-26 September 2012. Proceedings.

Omicini, A. (2007).
Formal ReSpecT in the A&A perspective.
Electronic Notes in Theoretical Computer Science, 175(2):97–117.
5th International Workshop on Foundations of Coordination Languages and Software Architectures (FOCLASA'06), CONCUR'06, Bonn, Germany, 31 August 2006. Post-proceedings.

Omicini, A. (2011).
Self-organising knowledge-intensive workspaces.
In Ferscha, A., editor, *Pervasive Adaptation. The Next Generation Pervasive Computing Research Agenda*, chapter VII: Human-Centric Adaptation, pages 71–72. Institute for Pervasive Computing, Johannes Kepler University Linz, Austria.

References IV

Omicini, A. and Denti, E. (2001).
From tuple spaces to tuple centres.
Science of Computer Programming, 41(3):277–294.

Omicini, A. and Viroli, M. (2011).
Coordination models and languages: From parallel computing to self-organisation.
The Knowledge Engineering Review, 26(1):53–59.
Special Issue 01 (25th Anniversary Issue).

Omicini, A. and Zambonelli, F. (1999).
Coordination for Internet application development.
Autonomous Agents and Multi-Agent Systems, 2(3):251–269.
Special Issue: Coordination Mechanisms for Web Agents.

Viroli, M. and Casadei, M. (2009).
Biochemical tuple spaces for self-organising coordination.
In Field, J. and Vasconcelos, V. T., editors, *Coordination Languages and Models*, volume 5521 of *LNCS*, pages 143–162. Springer, Lisbon, Portugal.
11th International Conference (COORDINATION 2009), Lisbon, Portugal, June 2009. Proceedings.

References V

Weyns, D., Omicini, A., and Odell, J. J. (2007).
Environment as a first-class abstraction in multi-agent systems.
Autonomous Agents and Multi-Agent Systems, 14(1):5–30.
Special Issue on Environments for Multi-agent Systems.

Zambonelli, F., Castelli, G., Ferrari, L., Mamei, M., Rosi, A., Di Marzo, G., Risoldi, M., Tchao, A.-E., Dobson, S., Stevenson, G., Ye, Y., Nardini, E., Omicini, A., Montagna, S., Viroli, M., Ferscha, A., Maschek, S., and Wally, B. (2011).
Self-aware pervasive service ecosystems.
Procedia Computer Science, 7:197–199.
Proceedings of the 2nd European Future Technologies Conference and Exhibition 2011 (FET 11).

$\mathcal{M}$olecules $\mathcal{of}$ $\mathcal{K}$nowledge

Self-Organisation in Knowledge-Intensive Environments

Stefano Mariani

s.mariani@unibo.it

DISI
Alma Mater Studiorum—Università di Bologna

Seminar @ DISI, Cesena - November 13th, 2014