

“Benchmarking via Challenges & Competitions”

... the RoCKIn legacy

... how I stopped worrying
about RoboCup and learned
to love benchmarking.

● ● ● *Matteo Matteucci, matteo.matteucci@polimi.it*

Dept. of Electronics, Information and Bioengineering

Politecnico di Milano

Where did it all start?

RoboCup European Championship Amsterdam-2000

- General Chair: Frans Groen, (groen@wins.uva.nl)
- Organizing Co-Chairs:
 - Enrico Pagello, (epv@dei.unipd.it)
 - Gerhard Kraetzschmar Ulm, (gkk@acm.org)
- Simulator League Chair: Huosheng Hu, (hhu@essex.ac.uk)
- F180 (Small Size) League Co-Chair: Andeas Birk, (cyrano@arti.vub.ac.be)
- F2000 (Middle Size) League Co-Chair: Pedro Lima, (pal@jsr.ist.utl.pt)
- Workshop Co-chair: Giovanni Adorni, (adorni@ce.unipr.it)
- Jarkko Kemppainen, (jarkko.kemppainen@nokia.com)
- Pierre Blazevic, (blazevic@robot.uvsq.fr)
- Andreas Zell, (zell@Informatik.Uni-Tuebingen.de)
- Raul Rojas, (rojas@inf.fu-berlin.de)
- Leo Stankevich, (stankevitch@phtf.stu.neva.ru)
- Josep De la Rosa, (pepluis@eia.udg.es)
- Silvia Coradeschi (silvia.coradeschi@tech.oru.se)
- **Advisory Board:**
 - Manuela Veloso, (mmv@school.coral.cs.cmu.edu)
 - Peter Stone, (pstone@research.att.com)
 - Hiroaki Kitano, (kitano@symbio.jst.go.jp)
 - Minoru Asada, (asada@er.ams.eng.osaka-u.ac.jp)



Amsterdam
2000

Where did it all start?

Bremen 2006



Competition as Experiments

“Challenge and competition events in robotics provide an excellent vehicle for advancing the state of the art and evaluating new algorithms and techniques in the context of a common problem domain. [...] treat competitions and challenges as repeatable experiments.”

Monica Anderson, Odest Chadwicke Jenkins, and Sarah Osentoski
“Recasting Robotics Challenges as Experiments”

Competition as Experiments

“One-time demonstrations of robot performance (e.g., grand challenges or other competitions) in robotics are one way of comparing the performance of robots, but they do not necessarily prove that one’s robotics research is consistently better or worse than another lab’s.”

Leila Takayama (Google[x], formerly at Willow Garage)

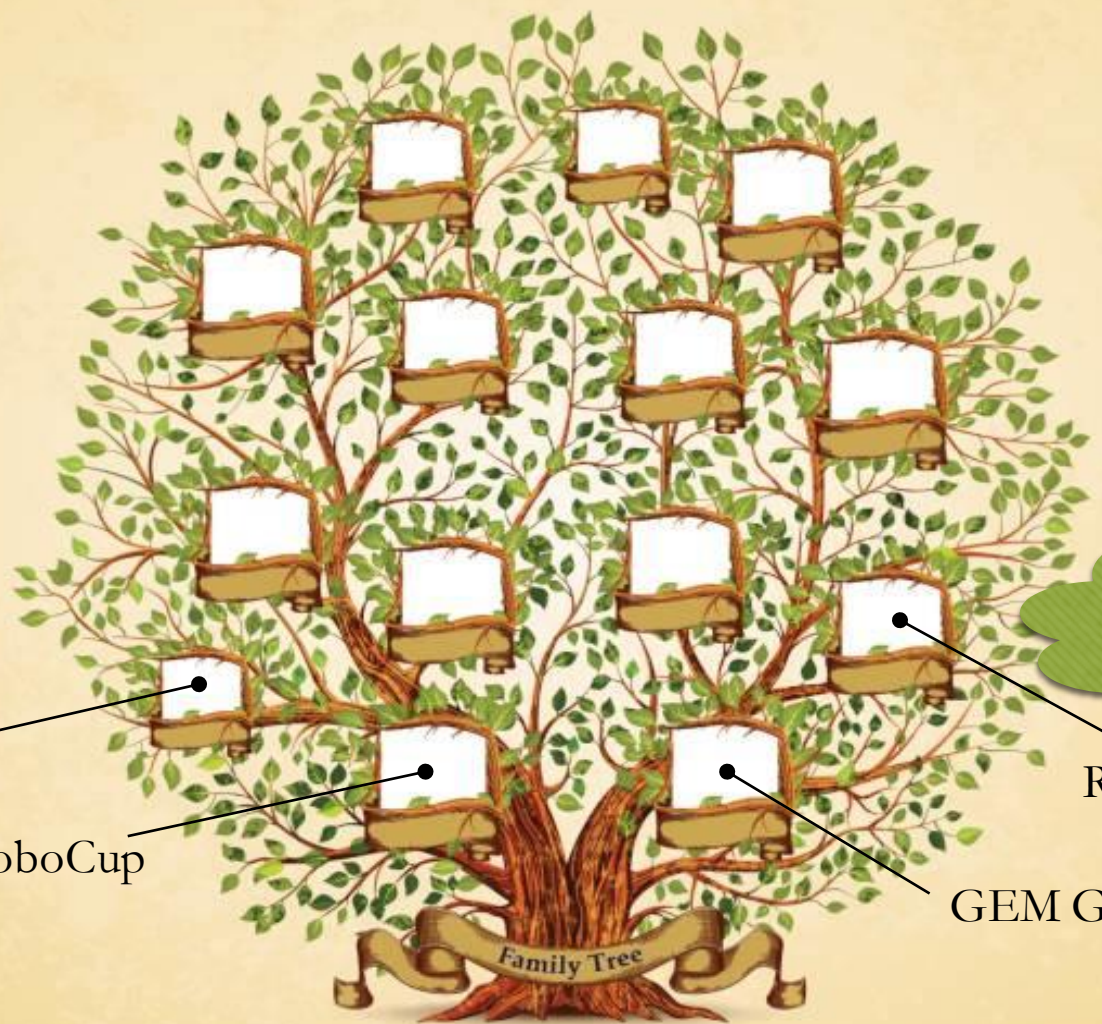
“Towards a Science of Robotics: Goals and Standards for Experimental Research”

Good Experimental Methodologies

- “General Guidelines for Robotics Papers using Experiments”

- Is it an experimental paper?
- Are the system assumptions/hypotheses clear?
- Are the performance criteria spelled out explicitly?
- What is being measured and how?
- Do the methods and measurements match the criteria?
- Is there enough information to reproduce the work?
- Do the results obtained give a fair and realistic picture of the system being studied?
- Are the drawn conclusions precise and valid





RoCKIn

RoboCup

RAWSEEDS

GEM Guidelines

I talked about this at the
CHIST-ERA Workshop
in Warsaw (2017)

The RoCKIn Idea

Robot Competitions Kick Innovation
in Cognitive Systems and Robotics



The RoCKIn Project


©2013 RoCKIn project, contract no. FP7-ICT-601012

Pedro U. Lima (IST-ID), Daniele Nardi (UNIROMA1),
Gerhard Kraetschmar (BRSU), Rainer Bischoff (KUKA),
Matteo Matteucci (POLIMI), Graham Buchanan (INNO)

The RoCKIn idea

- Build upon the well-established infrastructure of **RoboCup** competitions **plus**:
 - Introducing the **networked robot systems** (multiple robots, multiple sensors and devices in the environment)
 - adding further **natural interaction** between robots and humans, based on cognitive systems and principles
 - reducing the number and importance of subjective evaluation vs **objective evaluation in the competitions**
 - revising the evaluation criteria of tests such that the overall measure combines the **quality of subsystems** and the success in performing the **overall task**
 - applying more care in the design and revision of the competition rules for **better comparison across years**
 - **lowering the entry barrier for new teams**

RoCKIn project, contract no. FP7-ICT-601012



“Robots are doomed to close the loop”

- Robots experiments can be done at two different levels
 - **Task Level:** evaluation of whole systems on a specific task (e.g., the “bring me the glasses” tasks)
 - **Functionality Level:** evaluation of modules implementing, in a general manner, functionalities required by the competition tasks (e.g., vision, planning, manipulation)
- RoCKIn competitions should allow independent evaluation at both levels

Intelligence comes from
the effective interaction
of different capabilities

The RoCKIn Idea

Robot Competitions Kick Innovation
in Cognitive Systems and Robotics




The RoCKIn Project
©2013 RoCKIn project, contract no. FP7-ICT-601012

Pedro U. Lima (IST-ID), Daniele Nardi (UNIROMA1),
Gerhard Kraetschmar (BRSU), Rainer Bischoff (KUKA),
Matteo Matteucci (POLIMI), Graham Buchanan (INNO)

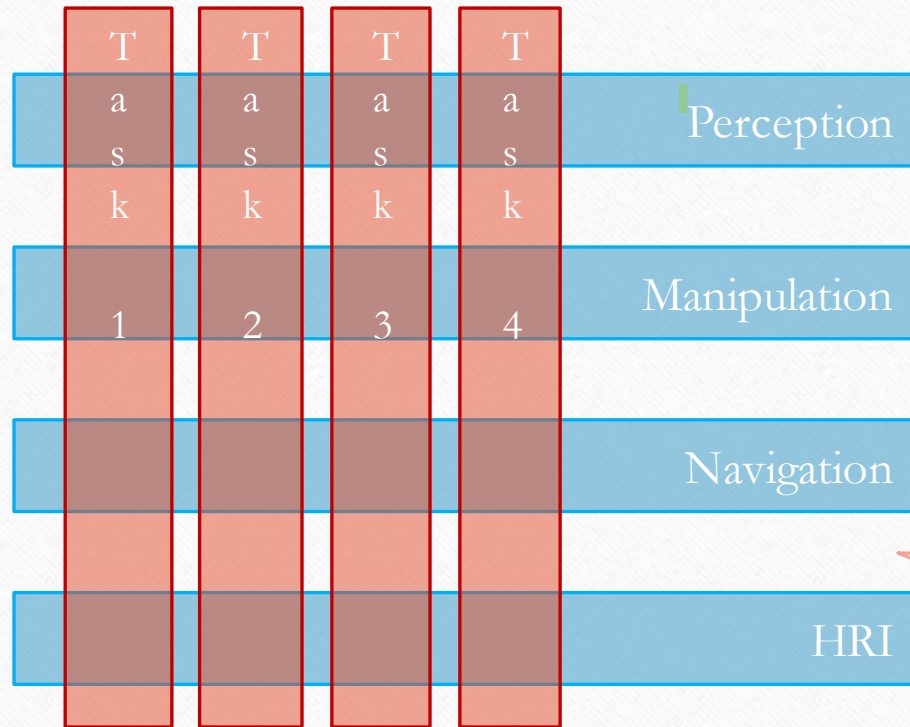
The RoCKIn idea

- Build upon the well-established infrastructure of **RoboCup** competitions **plus**:
 - Introducing the **networked robot systems** (multiple robots, multiple sensors and devices in the environment)
 - adding further **natural interaction** between robots and humans, based on cognitive systems and principles
 - reducing the number and importance of subjective evaluation vs **objective evaluation in the competitions**
 - revising the evaluation criteria of tests such that the overall measure combines the **quality of subsystems** and the success in performing the **overall task**
 - applying more care in the design and revision of the competition rules for **better comparison across years**
 - **lowering the entry barrier for new teams**

in project, contract no. FP7-ICT-601012



“Robots are doomed to close the loop”



Most of what we call today performance evaluation in intelligent systems is somehow limited to functionality benchmarks!

Functionality Benchmarks (FBM) evaluate performance of modules for specific functionalities

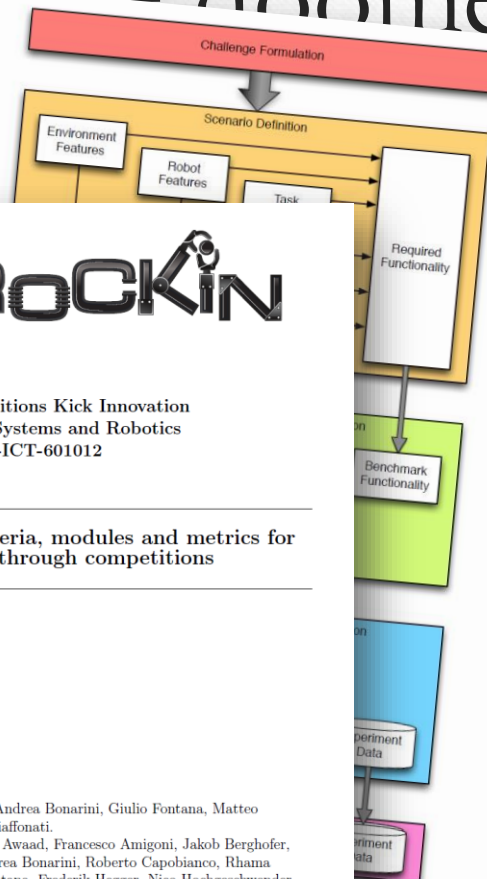
Task Benchmarks (TBM) assess performance of integrated robot systems facing complex tasks

“Robots are doomed to close the loop”

Most of what we call today competitions/challenges are task benchmarks!

Functionality Benchmarks (FBM) evaluate performance of modules for specific functionalities

Task Benchmarks (TBM) assess performance of integrated robot systems facing complex tasks



ROCKIN

SEVENTH FRAMEWORK PROGRAMME

RoCKIn - Robot Competitions Kick Innovation in Cognitive Systems and Robotics
FP7-ICT-601012

Grant Agreement Number: 601012
Funding Period: 01.01.2013 - 31.12.2014
Instrument: Coordination and Support Action


Specification of General Features for Benchmarking Tasks

Aamir Ahmad, Iman Awaad, Francesco Amigoni, Andrea Bonarini, Rhama Dwiputra, Giulio Fontana, Nico Hochgeschwender, Luca Iocchi, Gerhard Kraetzschmar, Daniele Nardi, Viola Schiaffonati, Sven Schneider, Matteo Matteucci

TÉCNICO LISBOA

KUKA

Lead Contractor for this Deliverable
Due date of deliverable: June 30, 2014
Actual submission date: July 10, 2014
Dissemination level: Public
Revision: 1.0

 **ROCKIN**
Funded by the European Union


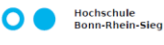



Robot Competitions Kick Innovation in Cognitive Systems and Robotics
FP7-ICT-601012

General evaluation criteria, modules and metrics for benchmarking through competitions

Deliverable: D-1.2
Due Date: June 30, 2014
Latest Update: July 10, 2014
Revision: 1.0

Editors: Francesco Amigoni, Andrea Bonarini, Giulio Fontana, Matteo Matteucci, Viola Schiaffonati.

Contributors: Aamir Ahmad, Iman Awaad, Francesco Amigoni, Jakob Berghofer, Rainer Bischoff, Andrea Bonarini, Roberto Capobianco, Rhama Dwiputra, Giulio Fontana, Frederik Hegger, Nico Hochgeschwender, Luca Iocchi, Gerhard Kraetzschmar, Pedro Lima, Matteo Matteucci, Daniele Nardi, Viola Schiaffonati, Sven Schneider.

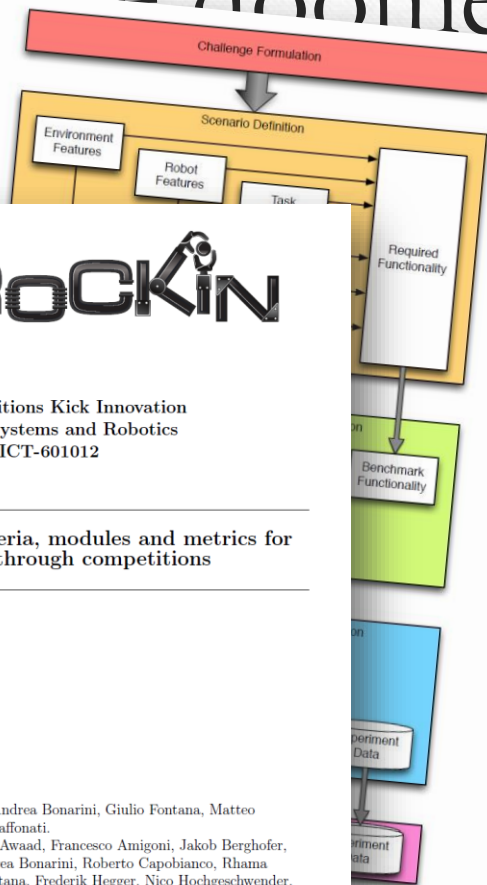
  
  

“Robots are doomed to close the loop”

This integrated approach is what the RoCKIn PO dubbed as «RoCKIn legacy» !!!

Functionality Benchmarks (FBM) evaluate performance of modules for specific functionalities

Task Benchmarks (TBM) assess performance of integrated robot systems facing complex tasks



 **ROCKIN**
Funded by the European Union



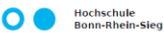



**Robot Competitions Kick Innovation
in Cognitive Systems and Robotics
FP7-ICT-601012**

**General evaluation criteria, modules and metrics for
benchmarking through competitions**

Deliverable: D-1.2
Due Date: June 30, 2014
Latest Update: July 10, 2014
Revision: 1.0

Editors: Francesco Amigoni, Andrea Bonarini, Giulio Fontana, Matteo Matteucci, Viola Schiaffonati.

Contributors: Amir Ahmad, Iman Awaad, Francesco Amigoni, Jakob Berghofer, Rainer Bischoff, Andrea Bonarini, Roberto Capobianco, Rhama Dwiputra, Giulio Fontana, Frederik Hegger, Nico Hochgeschwender, Luca Iocchi, Gerhard Kraetzschmar, Pedro Lima, Matteo Matteucci, Daniele Nardi, Viola Schiaffonati, Sven Schneider.

 **ROCKIN**

**RoCKIn - Robot Competitions Kick Innovation
in Cognitive Systems and Robotics
FP7-ICT-601012**

Grant Agreement Number: 601012
Funding Period: 01.01.2013 - 31.12.2015
Instrument: Coordination and Support Action

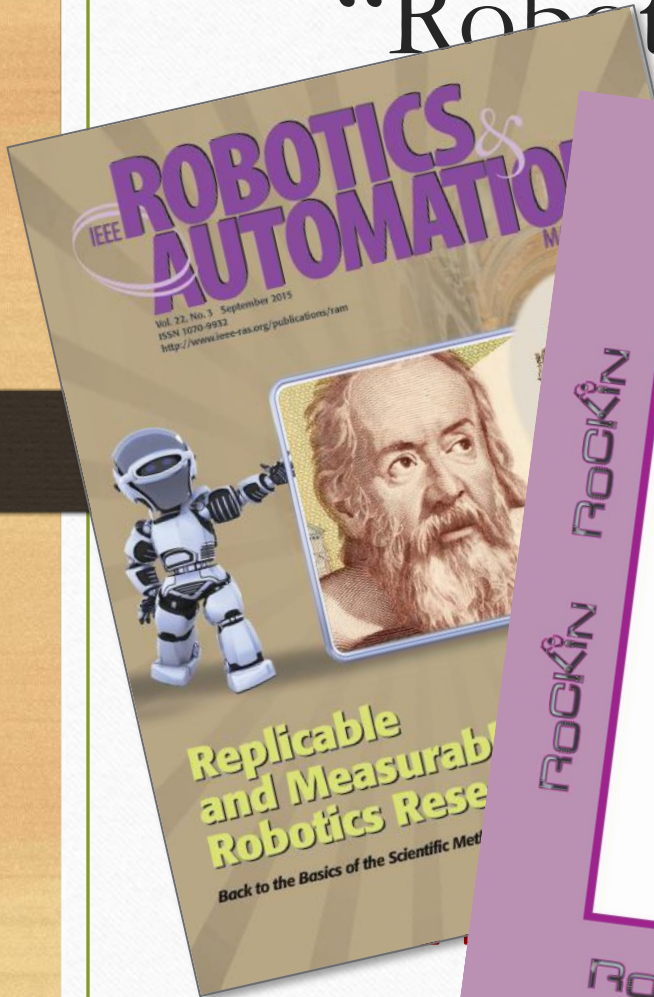
**Specification of General Features
and Metrics for Benchmarking Tasks
in Robot Competitions**

Aamir Ahmad, Iman Awaad, Francesco Amigoni,
Andrea Bonarini, Rhama Dwiputra, Giulio
Fontana, Nico Hochgeschwender, Luca Iocchi,
Gerhard Kraetzschmar, Pedro Lima, Matteo
Matteucci, Daniele Nardi, Viola Schiaffonati,
Sven Schneider

Lead Contractor for this Deliverable
Due date of deliverable:
Actual submission date:
Dissemination level:
Revision:

“Robots are doomed to close the loop”



ROCKIN ROCKIN

Competitions for Benchmarking

Task and Functionality Scoring Complete Performance Assessment

By Francesco Amigoni, Emanuele Bastianelli, Jakob Behringer, Andrea Bonarini, Giulio Fontana, Nico Hochgeschwender, Luca Iocchi, Gerhard K. Kraetschmar, Pedro Lima, Matteo Matteucci, Pedro Miraldo, Daniele Nardi, and Viola Schiaffonati

Scientific experiments and robotic competitions share some common traits that can put the debate about developing better experimental methodologies and replicability of results in robotics research on more solid ground. In this context, the Robot Competitions Kick Innovation in Cognitive Systems and Robotics (RoCKIn) project aims to develop competitions that come close to scientific experiments, providing an objective performance evaluation of robot systems under controlled and replicable conditions. In this article, by further articulating some results from the 2014 first RoCKIn competition, and by considering the RoCKIn approach offers tools that enable the replicability of experimental results.

Robotic Competitions and Challenges
Within the debate about the development of rigorous experimental methodologies in robotics research, the robotic competitions have emerged as a way to promote comparison of different algorithms

Original Content Identifier: 10.1109/RA.2015.2448871
Date of Publication: 11 September 2015

ROCKIN ROCKIN

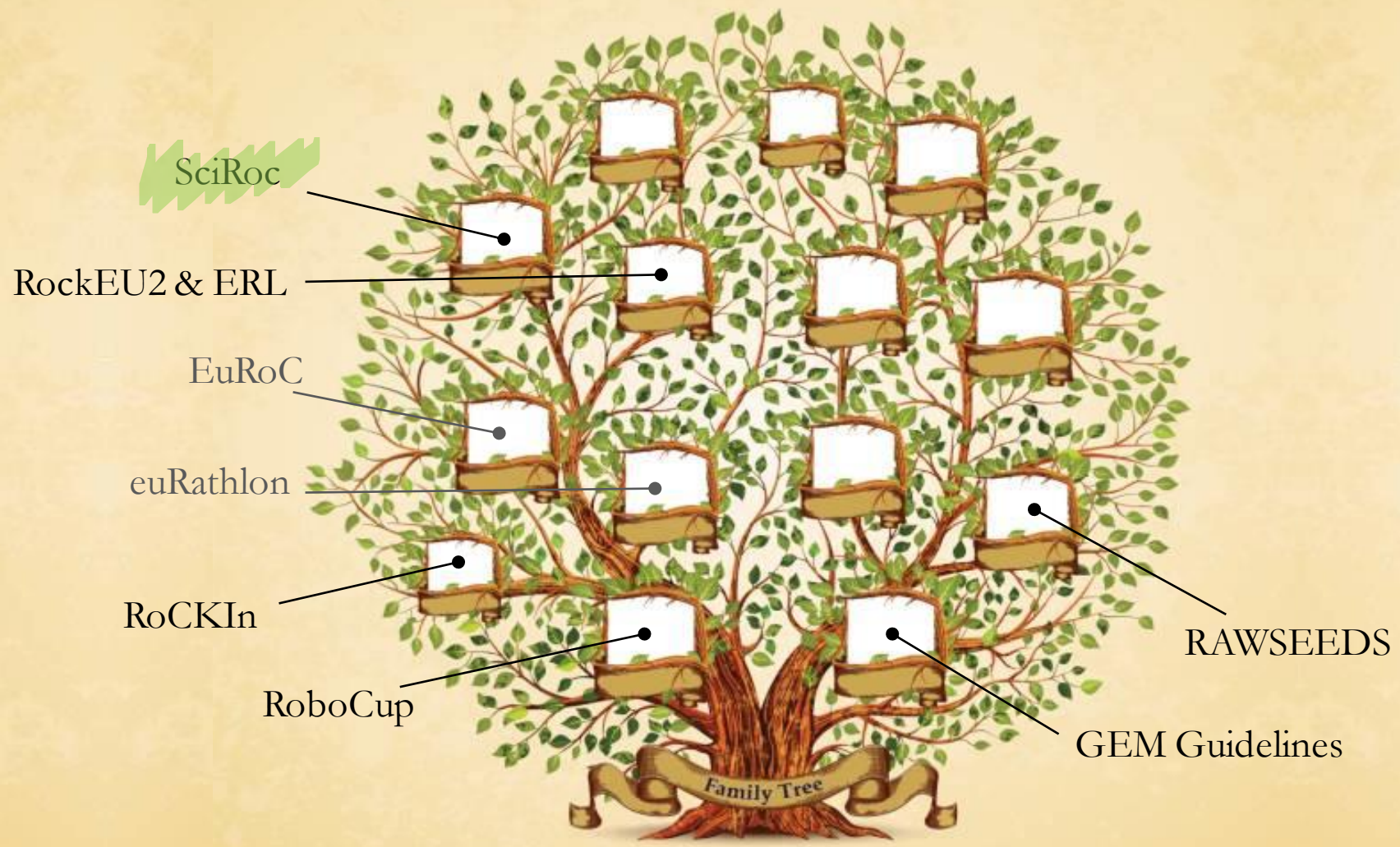
1000-9925/15/02015IEEE

Authorized licensed use limited to: Politecnico di Milano. Downloaded on September 24, 2020 at 23:02:07 UTC from IEEE Xplore. Restrictions apply.

This integrated approach is what the RoCKIn PO dubbed as «RoCKIn legacy» !!!

Functionality Benchmarks (FBM) evaluate performance of modules for specific functionalities

Task Benchmarks (TBM) assess performance of integrated robot systems facing complex tasks



The SciRoc Idea

EUROPEAN ROBOTICS LEAGUE

UWE Bristol

Hochschule Bonn-Rhein-Sieg University of Applied Sciences

ERL Consumer

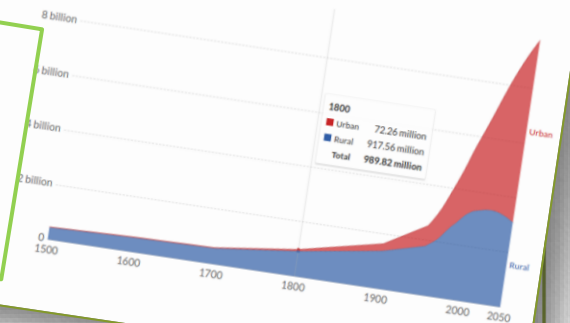
ERL Emergency

ERL Professional

ERL Smart Cities

Motivation and Background

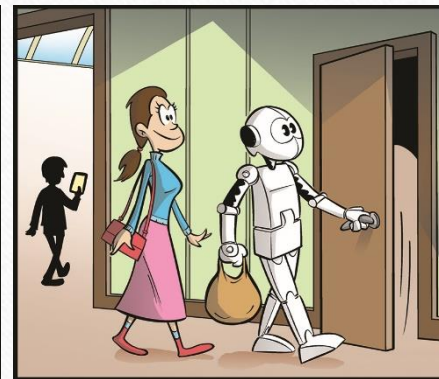
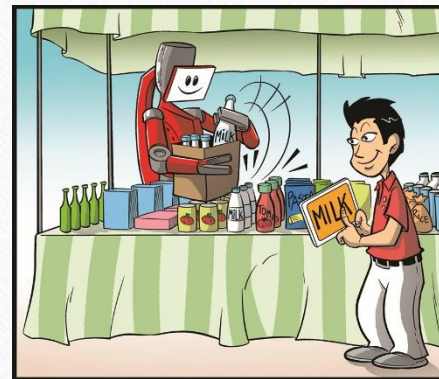
- By 2050, nearly 7 Billion people will live in cities
 - Most robots will be in cities too
- Smart Cities sense, process, and act
 - Offering data services and infrastructure to support robots
 - Robots and Smart Cities add value to each other
- The SciRoc Smart City Challenge
 - Brings robots to the city centre
 - Demonstrates the state of the art
 - Believable and Relatable
 - Stimulates discussion about our shared future



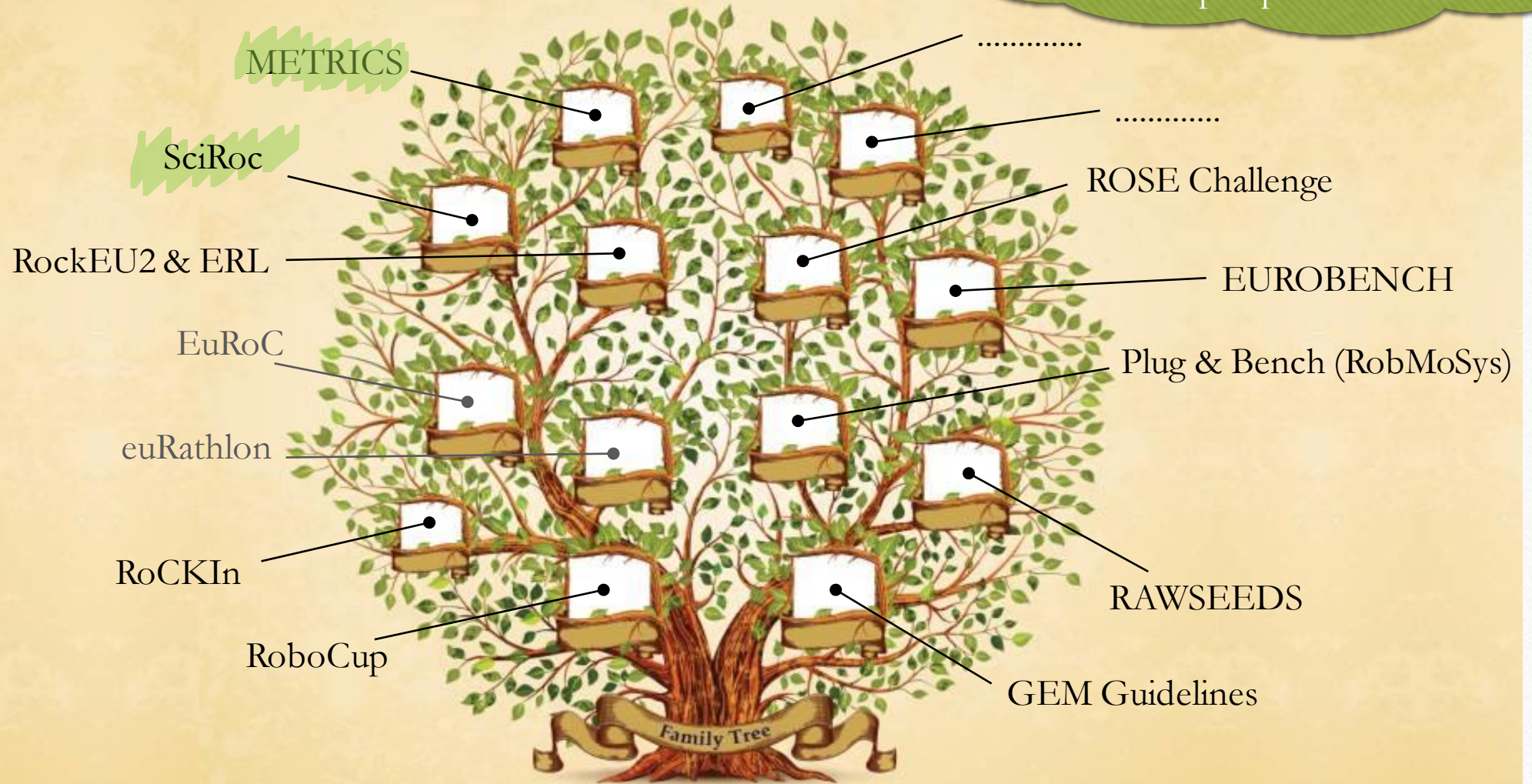
Sciroc Episodes

Human-robot interaction in social environments as well as social impact are the focus

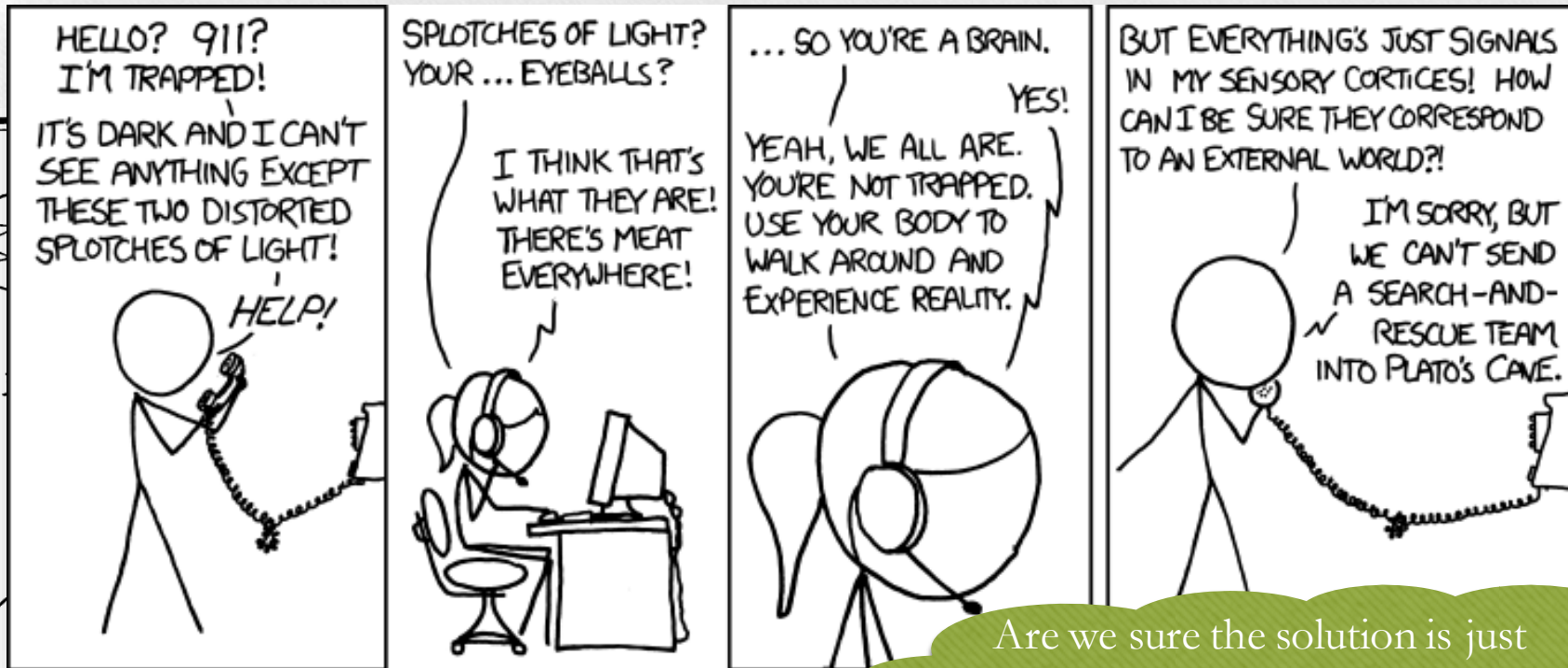
- Deliver coffee shop orders
- Talk the elevator
- Shopping pick and pack
- Open the door
- Fast delivery of emergency pills



Still lot to do from a
methodological/metrological
perspective ...



Back to the Body and Mind Problem ...



Are we sure the solution is just
to put the best AI on top of
the best mechanics?

“Benchmarking via Challenges & Competitions”
... the RoCKIn legacy

Thank you!

*Matteo Matteucci, matteo.matteucci@polimi.it
Dept. of Electronics, Information and Bioengineering
Politecnico di Milano*