INFOCOMP 2016 International Expert Panel:

Practical Experiences and Best Practice in Scientific and High End Computing. Is the Future Data-centric and Computing-centric?

May 26, 2016, Valencia, Spain

The Sixth International Conference on
Advanced Communications and Computation
(INFOCOMP 2016)

INFOCOMP
May 22–26, 2016 - Valencia, Spain

INFOCOMP Expert Panel: Is the Future Data-centric & Computing-centric?

Panelists

- *Claus-Peter Rückemann* (Moderator),
  Westfälische Wilhelms-Universität Münster (WWU) /
  Leibniz Universität Hannover /
  North-German Supercomputing Alliance (HLRN), Germany
- *Małgorzata Pankowska,*
  Department of Informatics, University of Economics in Katowice, Poland
- *Iryna Lishchuk,*
  Institute for Legal Informatics, Leibniz Universität Hannover, Germany
- *Roderick Melnik,*
  Wilfrid Laurier University, Canada
- *Sandra Sendra Compte,*
  Signal Theory, Telematics and Communications Department (TSTC), University of Granada, Spain

INFOCOMP 2016: `http://www.iaria.org/conferences2016/INFOCOMP16.html`
Program: `http://www.iaria.org/conferences2016/ProgramINFOCOMP16.html`

INFOCOMP Expert Panel: Is the Future Data-centric & Computing-centric?

Panel Statements:

- **Practical Experiences:** Long-term multi-disciplinary data, High End Computing and storage, supercomputing, Big Data types / handling (Volume, Variability, Velocity, Vitality, Veracity), reasonable, reusable, portable, commonly available standards, technologies, and methods.
- **Best Practice:** Long-term essential content and context should preceed computational needs: Data and structure preceeds computation for long-term.
- **Scientific computing:** Scientific application scenarios with different requirements.
- **High End:** Limits of bandwidth and latency regarding transfer and storage (much more than computing).
- **Knowledge resources:** Conceptual knowledge, classification, managing complexity, ...
- **Data-centric:** Data handling priority. View of disciplines.
- **Computing-centric:** Computing priority. Resources providers' view.

INFOCOMP Expert Panel: Is the Future Data-centric & Computing-centric?

Pre-Discussion-Wrapup:

- **Focus:** Experiences with differences of data-centric / computing-centric?
- **Recommendations:** Which general long-term solutions and recommendations?
- **How-to:** How can sustainable 'centric' solutions be created?
- **Long-term:** Priorities and scenarios with centricites?
- **Context:** Are there integrated centricities?
- **Sustainability:** Scenarios beyond multi-disciplinary and long-term?
- **Networking:** Discussion! Open Questions? Suggestions for next Expert Panel?

INFOCOMP Expert Panel: Post-Panel-Discussion Summary

Post-Panel-Discussion Summary (2016-05-27):

- Data-centric, computing-centric, . . .: **Purpose** is important.
- **Advanced methods** required for a) factual, conceptual, procedural, metacognitive, . . . documentation, classification, concordances, b) ontologies, c) authentication / long-term signatures. **Fuzziness** required for documentation / description.
- **Data-centric / knowledge-centric**: Often better adopted to requirements of disciplines and content, management of complexity, data creation and documentation. Long-term scientific computing often benefits from data-centric approaches. Sustainability will significantly benefit.
- **Computing-centric**: Benefits for computing-intensive, often affords porting to architectures (resources life-time, access to high end resources, . . .).
- Documentation should be done **by owners/researchers**.
- **Strengthen the rights and possibilities of owners**. Handling originator/rights/legal/license issues seems feasible on long-term.
- Multi-disciplinary long-term solutions, infrastructures, and integrative facilities required. Example: Gather knowledge on volcanic eruptions from historical oil-paintings complementing present research (phenomena, locations, visibility, climate, . . .).
- Funding agencies should update their **best practice**: Make "long-term data obligations" for funded research.
- **Foster awareness** of data value: From funding, politics, researchers – everyone.
- The earlier to realise the better.

INFOCOMP Expert Panel: Table of Presentations, Attached

Panelist Presentations: (presentation order, following pages)

- **Data Should be Value, Computing Should be the Way, and Experiences and Best Practice Should be the A and O.** (*Rückemann*)
- **Is the future data-centric and computing-centric?** (*Pankowska*)
- **In the era of large scale data and large scale computing the focus is to integrate old and new data into validated information and to convert this information into knowledge directly applicable for use** (*Lishchuk*)
- **Scientific computing for multiscale problems, focusing on coupled phenomena at the nanoscale.** (*Melnik*)
- **Spontaneous Networks: The best way of obtaining computing resources and processing our information** (*Sendra Compte*)

INFOCOMP 2016 International Expert Panel:
Practical Experiences and Best Practice in Scientific and High End Computing.
Is the Future Data-centric and Computing-centric?

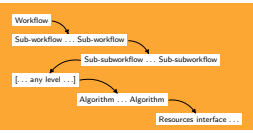
Data Should be Value, Computing should be the Way, and Experiences and Best Practice should be the A and O

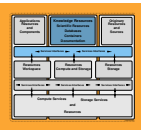
The Sixth International Conference on Advanced Communications and Computation
(INFOCOMP 2016)
May 26, 2016, Valencia, Spain

Dr. rer. nat. Claus-Peter Rückemann[1,2,3]

[1] Westfälische Wilhelms-Universität Münster (WWU), Münster, Germany
[2] Leibniz Universität Hannover, Hannover, Germany
[3] North-German Supercomputing Alliance (HLRN), Germany

ruckema(at)uni-muenster.de



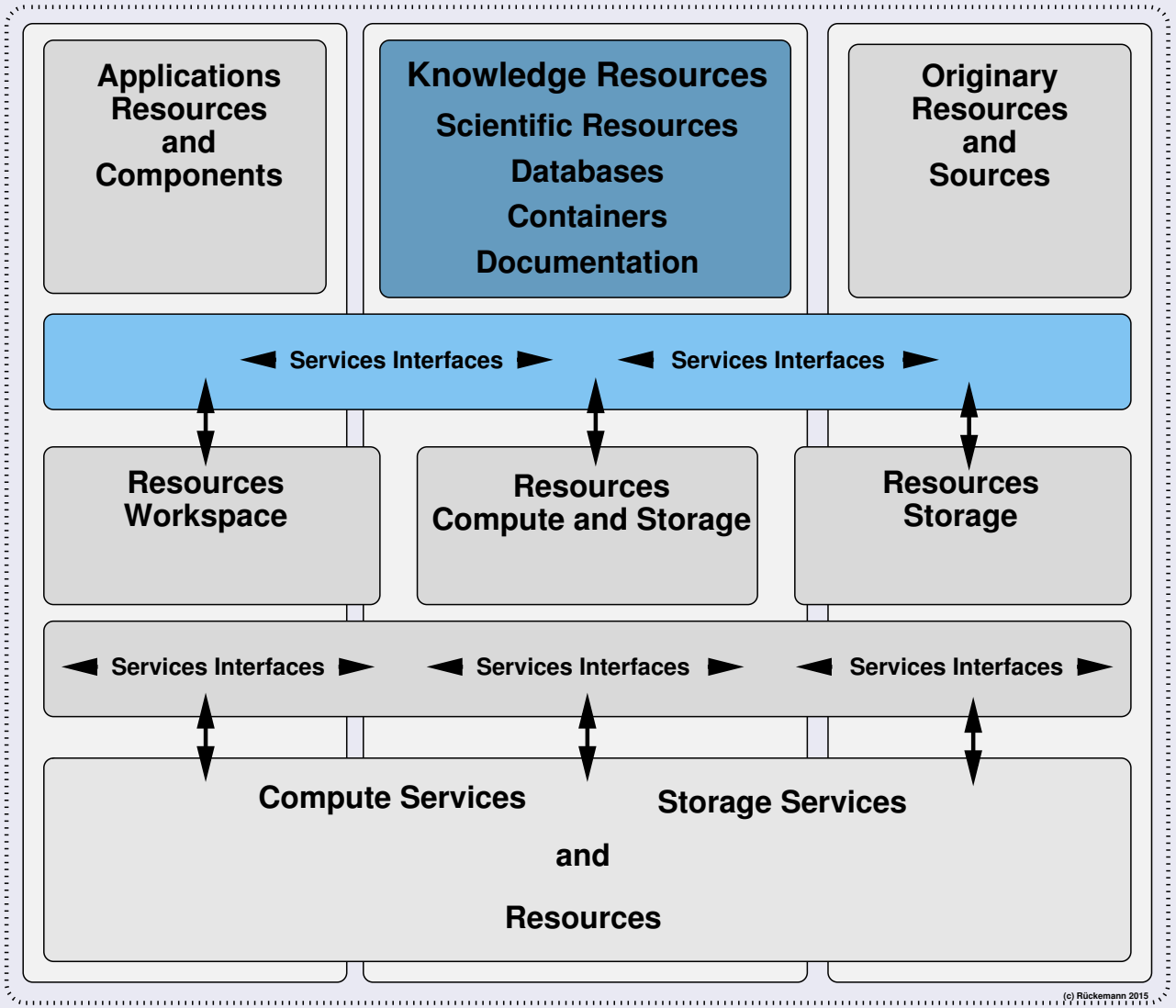

Data-centric / computing-centric

- **Scenarios:** Most resources providers favor computing-centric, e.g., for simple application on High End Computing resources.
- **Data esteem:** Data-centric for data esteem.
- **Applications:** Computing-centric for applications and commerce.
- **Integration:** Computing-centric data should be integrated into data-centric world.
- **Documentation:** ... major importance with research data management.
- **Content and applications:** Natural sciences/fundamental research, applied sciences/practical applications,
  ... are not long-term integrated in theory and practice.
- **Monolithic architectures:** System components require continuous re-development.
- **Big Data:** Classical methods (e.g., relational and object oriented) can hardly provide universaly efficient solutions.
- **Data size:** For decades, disk speeds and sizes do not keep up pace with data generation.
- **Knowledge:** Content and context are not appropriatley documented for decades.
- **Automation:** Instructive documentation is not available.

Missing and emerging solutions regarding content and scenarios

- **Content and applications:** Frameworks for long-term integration of fundamental research, content, practical applications.
- **Monolithic architectures:** Re-use design and implementation of components.
- **Big Data:** Advanced methods and new algorithms (e.g., NoSQL).
- **Data size:** 1: Increase of reliable and cheap disk/storage speeds and sizes; 2: Reduce data sizes.
- **Knowledge:** Knowledge-based documentation of content and context (e.g., knowledge resources).
- **Automation:** Instructive documentation for knowledge.
- **Transfer and storage:** Significantly (on-demand) increase bandwidths, decrease latencies.
- **Value:** Support the value of data and knowledge with best practice and funding (creation, documentation, computing, storage, integration, ...).
- **Standards:** Modularise standards' integration, support long-term standards.
- **Resources:** Modularise complex-to-manage high end HW and SW, empower users to handle technology, reduce costs of lifecycles and energy consumption.
- **Measurements and means:** Knowledge resources, e.g., long-term research data management/libraries.

Example Framework – Disciplines, Services, Providers



Integration & development of long-term knowledge & measurements

- **Solutions, which can be integrated.**
- **Improved data organisation, long-term data, structures, means.**
- **Knowledge documentation, content / context vitality.**
- **Creation of standards/systematics/methodologies with content.**
- **Long-term sustainability of universal knowledge discovery.**
- **Multi- and trans-disciplinary work.**
- **Support High End Computing, intelligent systems, education.**
- **Integrated Information and Computing System components.**
- **Mandatory best practice (e.g., for participation and funding).**

INFOCOMP 2016
Practical Experiences and Best Practice in Scientific and High End Computing

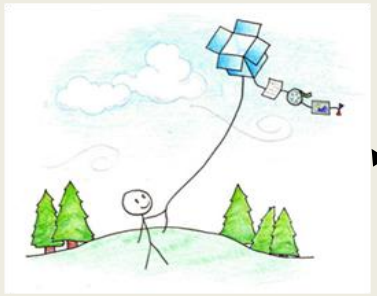
Is the future data-centric and computing-centric?

Malgorzata Pankowska
Department of Informatics
University of Economics in Katowice,
Poland

Is the future data-centric AND computing-centric?

- Is the future EITHER data-centric OR computing-centric?
- Is the future data-centric FOR computing-centric?
- Is the future computing-centric FOR data-centric?

Is the future data-centric FOR computing-centric



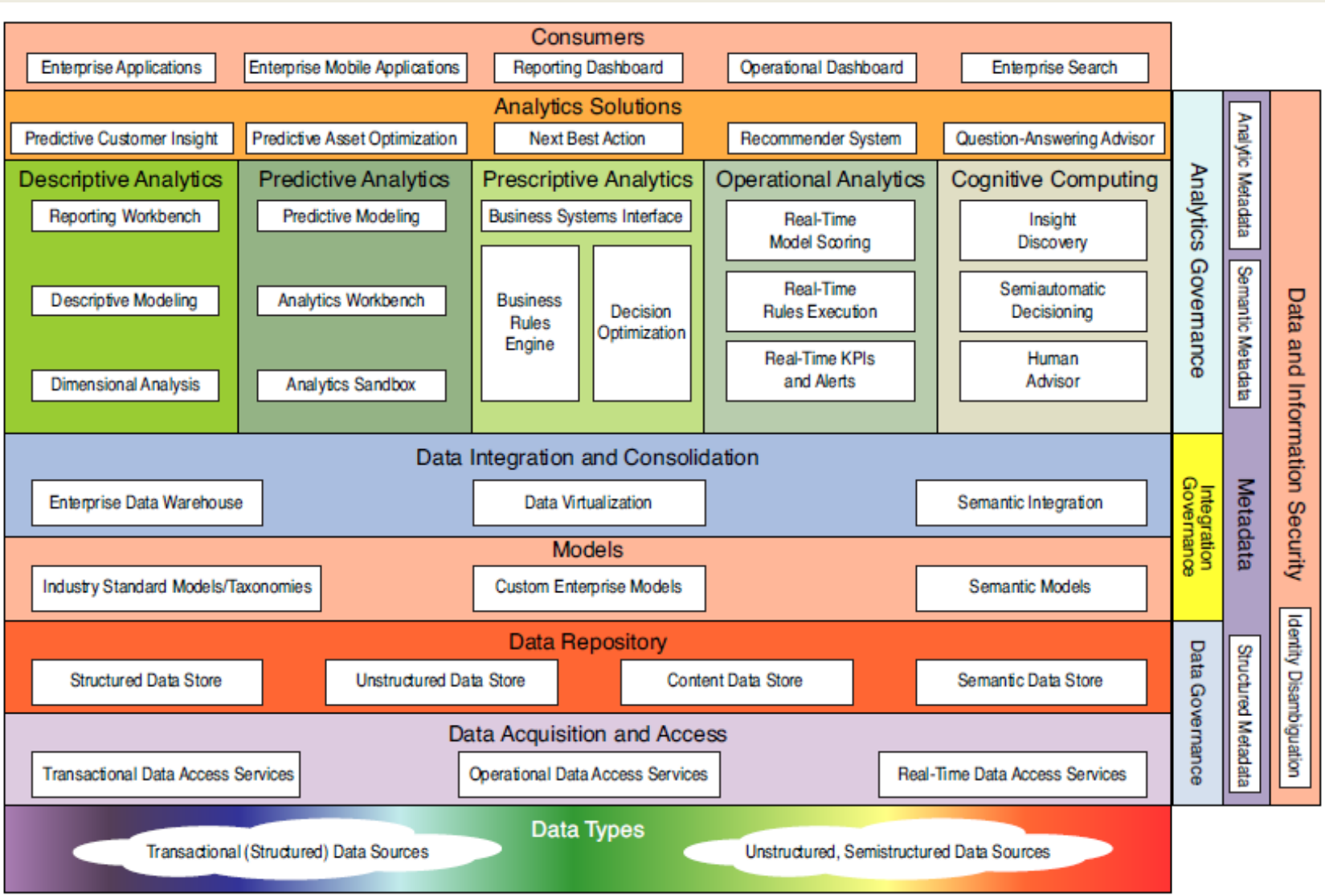
Is the future data-centric FOR computing-centric?

- Web 1.0 – the WWW evolution first stage
- Web 2.0 – Wiki-based platforms, blogs and microblogs, instant messaging programs, vide/photo sharing systems, and social/professional networks
- Web 3.0 – semantic Web,
- etc.

Event Stream Processing

- processing data continuously, on the move, in memory with very speed and low latency, parallel streams of data, application of rules and analysis using dataflow centric Event Stream processing (SAS ESP Studio)
- stream analytics for predictive asset maintenance
- sensors on transport means, car, trains, but also on medical devices,
- sensors to detect patterns in data or just situational monitoring,
- over 2.1 milion sensors generating 3 trillion rows of data/minute,
- contextual realtime analysis of the streaming phone data, real time decision manager,
- segmentation, sentiment analyses, text analyses, reliability analyses, predictive modelling

Architecture Building Bocks [Mitra, 2016]



System Autopoiesis

- **Ability of a system to generate its specific constitution – its components (structure) and their interplay (organization) – on its own [Yolles, 1999].**
- **The idea of autopoietic systems stems from the theory of social systems understood as systems of communication that reproduce all their necessary, specific structures within their own self-referential closed processes [Schumacher, 2011].**
-
- **Autopoietic systems are self-producing in a sense that they produce a network of processes that enables them to produce their own components**
-
- **Self-organization means that low-level interactions between individuals spontaneously emerge in certain properties. The properties, which are achieved through selfish actions of individuals, have certain functionality i.e., fulfill a purpose beneficial for the system as a whole.**
- **Self-organization can be defined as the emergence of coherent, global behaviour out of local interactions between components**

According to Whitaker [2011] a self-organization refers to a variety of distinct attributes such as:

- **self-creation:**
  - **system's origin is somehow determined by its character or the specific circumstances in which it occurs;**
- **self-configuration**
- **self-regulation:**
  - **system actively controls the course of its internal transformations;**
- **self-steering:**
  - **system actively controls its course of activity within some external environments;**
- **self-maintenance:**
  - **system preserves itself, its form and its functional status over time;**
- **self-production**
- **self-reference**

Distributed Computing Models properties:

- **self-configuration:**
  - the entities of systems can automate system configuration following high level specification, and can self-organize themselves into desirable structure and patterns, automatically configuring components to adapt them to different environments;
- **self-optimization:**
  - the system parts constantly seek improvement to their performance and efficiency, and are able to adapt to changing environment without direct human input**;**
- **self-healing:**
  - **the parts can automatically detect and recover from faults;**
- **self-protection:**
  - the system parts can automatically defend against attacks or isolate the attackers [Boutaba & Xiao, 2010, Wang et al., 2007]

Panel on INFOCOMP / MODOPT / SPWID
Topic: Practical Experiences and Best Practice in Scientific and High-end Computing; Is the Future Data-centric and Computing-centric?

Institut für Rechtsinformatik
Leibniz Universität Hannover/ Iryna Lishchuk, LL.M

CHIC
Computational Horizons in Cancer

Large Scale Computation

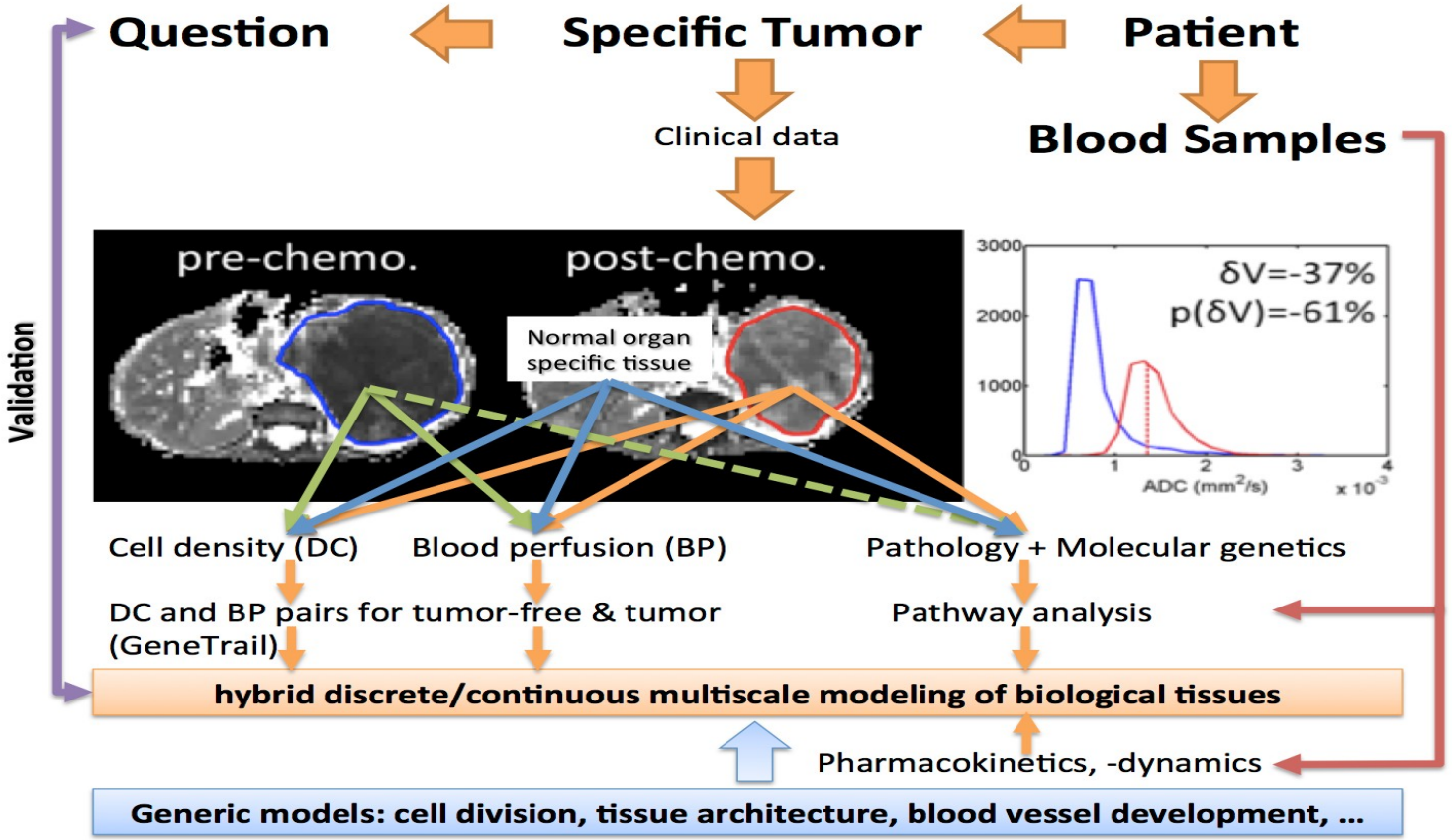
- COMPUTATIONAL MODELING
- The power of computational modeling is that it allows scientists and engineers to simulate variations more efficiently by computer, saving time, money, and materials.

- LARGE SCALE DATA
- Since we are moving into the era of the $1,000 genome analysis and the reality of mandatory EHRs, the focus of bioinformatics has shifted from gathering data to analyzing the massive amounts of available data for direct application in patient care.

Prof. Piotr Czauderna,

"Oncology, Age of Precision: Medicine and IT - clinician's perspective"

Computation Models in Oncology



yper-
Multiscale Models and Repositories for In Silico Oncology"
http://chic-vph.eu/project/

CHIC
Computational Horizons in Cancer

Large Scale Data in Healthcare

- By the usage of machine learning methods the accurate prediction of the hypermodels will improve over time.
- One can compare the output of the hypermodels of today with the weather forecast of 30 years ago.
- To use the result of a hypermodel for clinical decision support they need to be validated and it needs to be shown that decision based on prediction of a hypermodel is better than a decision based on the clinical standards of today.

CHIC Deliverable No. 2.5, "Clinical relevance of the CHIC project – Describing the integrated workflows of the scenarios from a clinical perspective

Liability Issues in Large Scale Computing

Liability, reliability, accuracy of IT solutions

Velocity, variety, veracity of data

I. Defects in software

II. Defects in algorithms, processing and compiling the data from different sources

III. Defects in data itself

Who is liable for what?

Defects in code, algorithms, data

- Software:
  - Each technical malfunction of software is a technical defect
  - *"Software is never bug-free"*
  - Fully functional code may be defective in legal sense
- Data:
  - The data is of under-mediate level if it substantially misrepresents the average value and places the prognosis decision on insecure basis
- Defects in data or low fault tolerance of software assessing the data or wrong algorithm in assessment?

Privacy Issues

- Article 9 (1) GDPR*
- processing *"data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation be prohibited"*,
  ….unless exceptions apply

* *REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)*

Legitimate Data Processing

- Article 9 (2) (j) GDPR
- "*processing is necessary for scientific …research purposes or statistical purposes in accordance with Article 89(1) based on Union or Member State law which shall be proportionate to the aim pursued, respect the essence of the right to data protection and provide for suitable and specific measures to safeguard the fundamental rights and the interests of the data subject.*"
- Article 6 (1) (a)
- Prior informed consent of the data subject

Privacy Issues

- Article 89 (1) GDPR
- *"appropriate safeguards…. shall ensure that technical and organisational measures are in place …. to ensure respect for the principle of data minimisation. Those measures may include pseudonymisation provided that those purposes can be fulfilled in that manner. Where those purposes can be fulfilled by further processing which does not permit or no longer permits the identification of data subjects, those purposes shall be fulfilled in that manner."*

IT Applications in Healthcare

"In the era of large scale data and large scale computing the focus is to integrate old and new data into validated information and to convert this information into knowledge directly applicable for use"

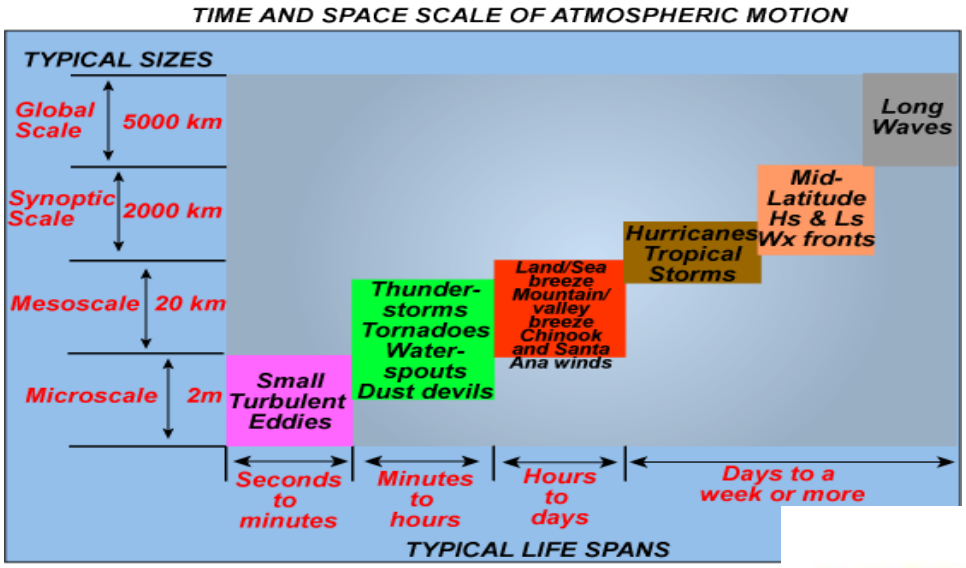
CHIC
Computational Horizons in Cancer

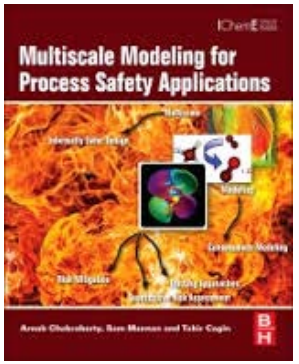
THANK YOU!

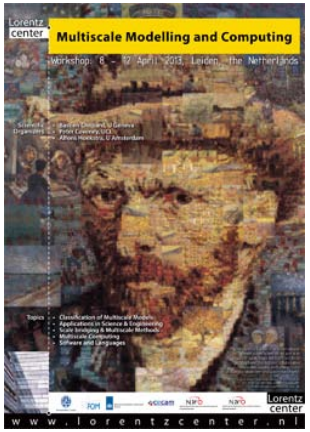
The research leading to these results has received funding from the European Union Seventh Framework Programme FP7/2007-2013 under grant agreement No 600841.

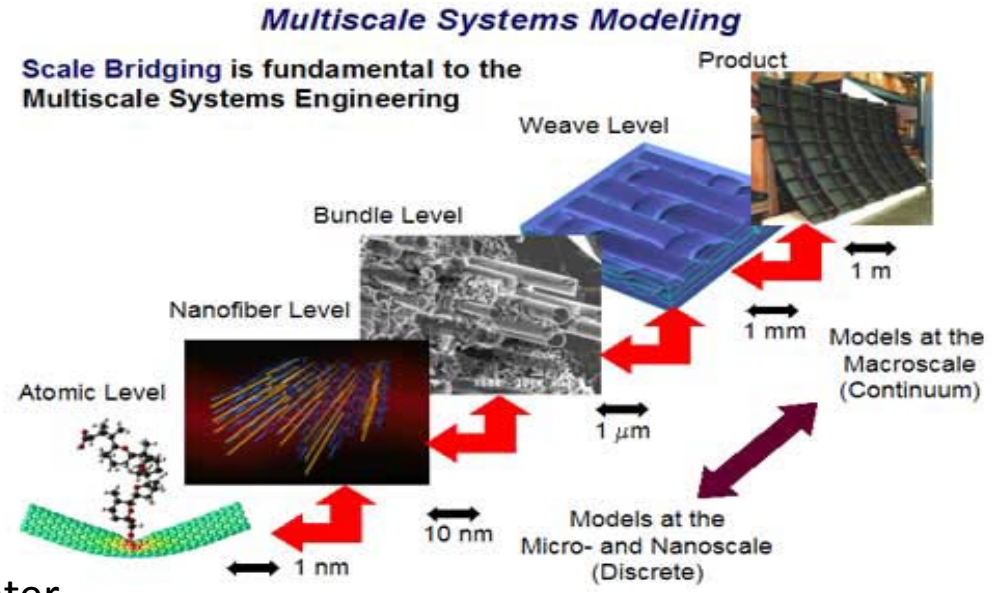
Roderick Melnik

Interacting Spatio-Temporal Scales: Going Up and Going Down





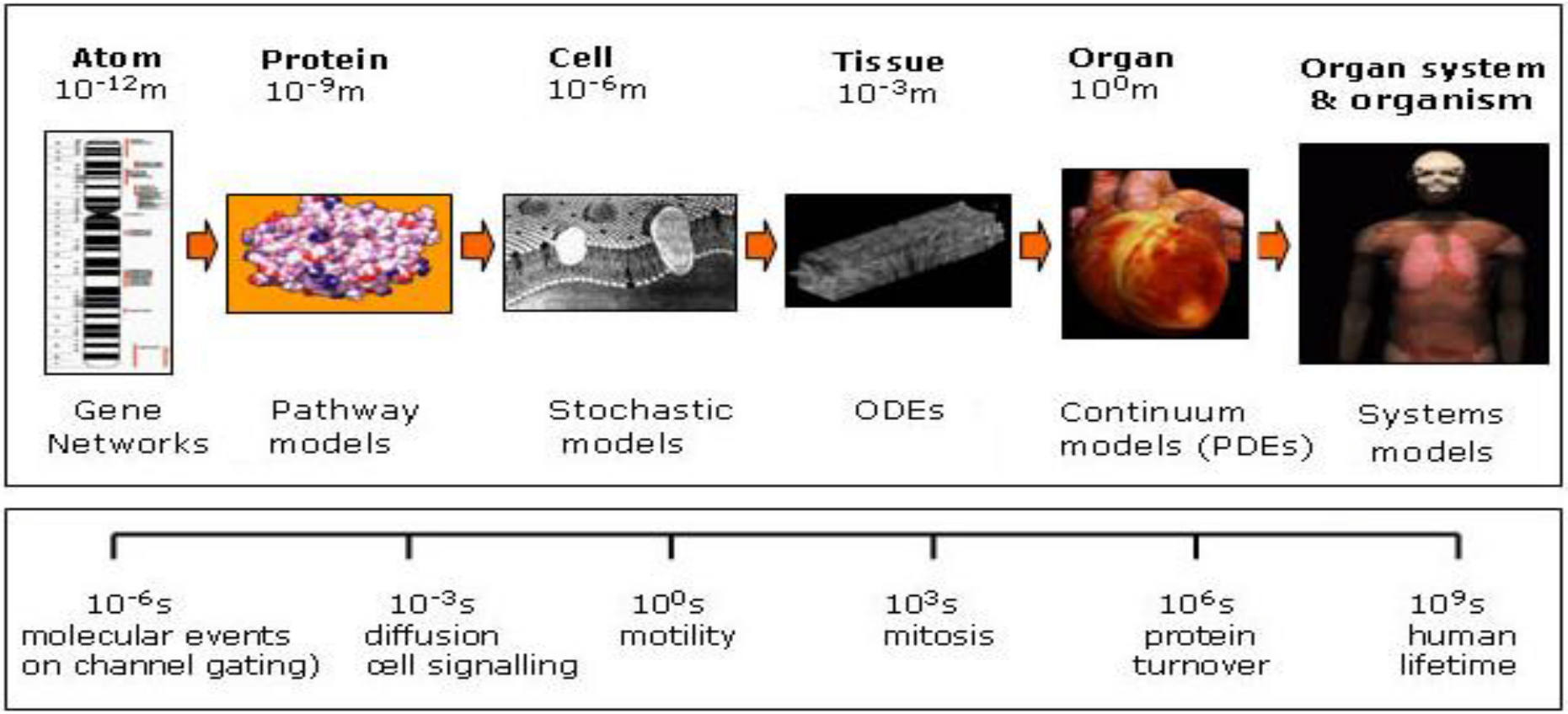




Acknowledge: BYU, UChicago, Lorentz Center

Importance and Universality of Interacting Spatio-Temporal Scales

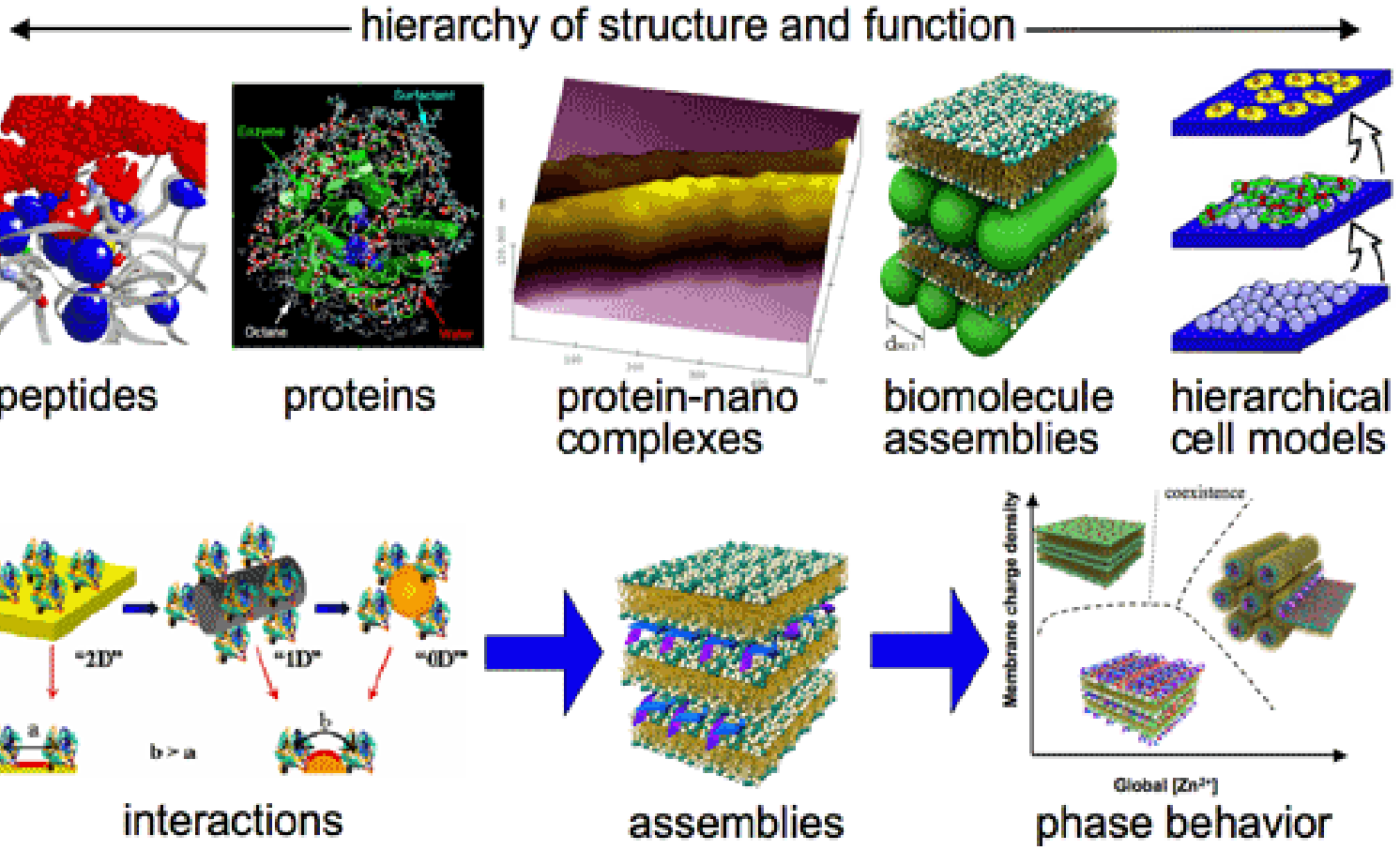
γνῶθι σεαυτόν → *Nosce te ipsum* → Know thyself



Biological Nanostructures and the Design of Life

1 nanometer = 10^{-9} meter

It is roughly 100,000 times smaller than the thickness of human hair.



Panel on INFOCOMP / MODOPT / SPWID

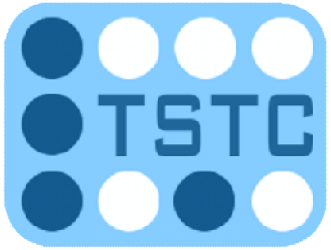
Practical Experiences and Best Practice in Scientific and High-end Computing; Is the Future Data-centric and Computing-centric?

Spontaneous Networks: The best way of obtaining computing resources and processing our information

Thursday, 26 May 2016
Valencia, Spain

Sandra Sendra Compte
(ssendra@ugr.es)

ugr | Universidad de Granada



Spontaneous Networks: The best way of obtaining computing resources and processing our information

ugr Universidad de Granada

Introduction

- The rapid development of the processing and storage technologies with the appearance of many Internet services has brought us cheaper, more powerful and more accessible computing resources.
- Cloud computing provides several features that make it attractive to business owners:
  - No up-front investment,
  - Decreasing the operating cost as well as reducing business risks and maintenance expenses
  - The networks become in scalable architectures with easy access.
- The evolution of smart mobile devices has facilitated to go a step beyond the simple cloud computing network.
- This is the creation of spontaneous mobile ad hoc cloud computing networks in which a mobile device connected to the cloud computing and belonging to the spontaneous ad hoc network let other mobile device to access to the computing resources to make some tasks.
- The main features of spontaneous networks are the following:
  - Network boundaries are poorly defined.
  - The network is not planned.
  - Hosts are not preconfigured.
  - There are not any central servers.
  - Users are not experts.

Definitions

Definitions:

- A spontaneous ad hoc network is a type of ad hoc network that is formed during a certain period of time, with no dependence on a central server and without the intervention of an expert user.
- This network is made of several independent nodes which are in the same place at the same time in order to communicate with each other
- They are basically those who seek to imitate human relationships in order to work together in groups, running on the already existing technology.
- Their objective is the integration of services and devices in an environment which allows the provision to the user of an instant service with minimum manual intervention, ensuring important aspects, such as the multimedia quality or network lifetime.

Network Model and Description

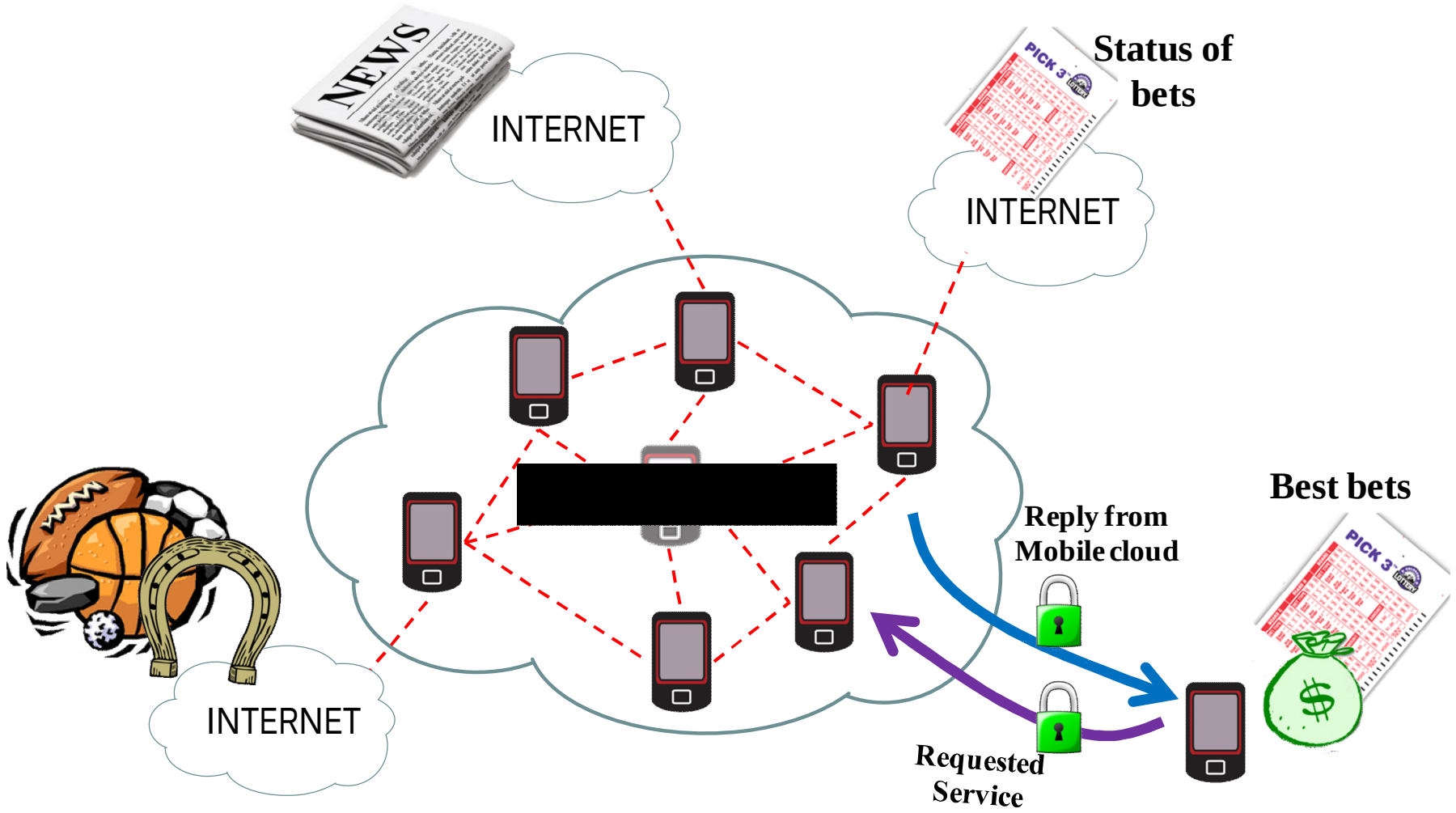
Network Model and Description:

Our network model meets the following requirements.

- Devices can move freely in the given area. Even out of each other's range.
- Every node is also a router. It has a limited communication range towards other nodes.
- The different identities are given by IP addresses. Each address is obtained dynamically following our previous works [1,2].
- There is no central administration.
- Devices can come from everywhere and join and leave at will.
- Resources for cloud computing can be provided by any node if it has enough capacity to do it.

Spontaneous Networks: The best way of obtaining computing resources and processing our information

Network Model and Description



Problems and Solutions

Problems:

- The biggest problem in these networks is the **security issue**.
- The use of a certificate authority (CA) server is not a good idea because of the lack of a robust infrastructure and the distance. The device could be very far from the CA, so their connection could be a big issue.
- In mobile computing, there are some inherent problems such as resource scarcity, frequent disconnections, and mobility that make exploiting its full potential difficult.
- The development of our secure application enables us to maintain the network anonymity and the delivered information.

Spontaneous Networks: The best way of obtaining computing resources and processing our information

Problems and Solutions

Solutions:

- Trusted network based on human factors.
- Distributed computation of security and computation issues among all network nodes. It included cloud connection and messages exchanges.
- Well-balance security overload. Using cryptographic algorithms (as other protocols) and checking operations on a random way to guarantee security and to avoid overload.

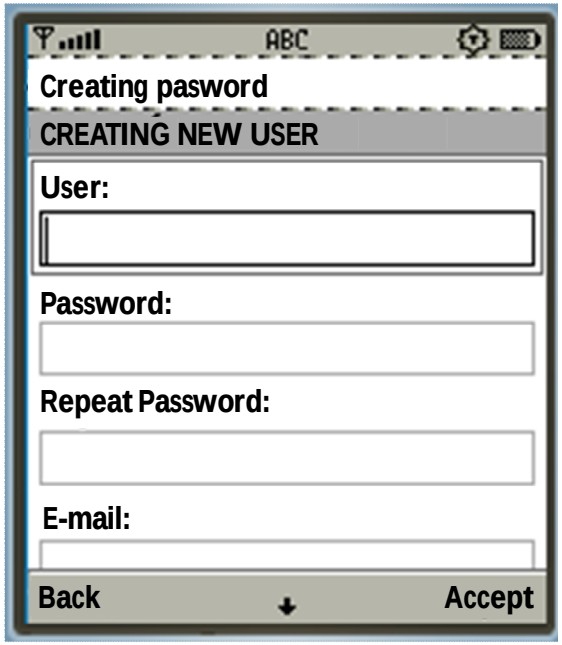
Our System:

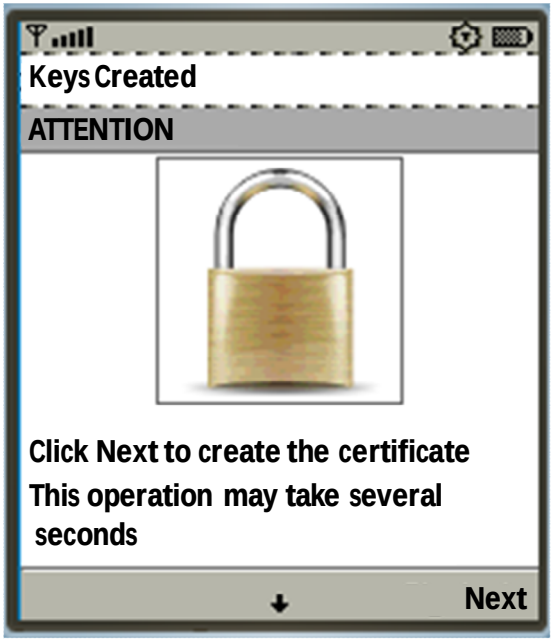
- Hierarchical system composed by two levels. The lowest level is the network level and it is in charge of providing security to the network. The upper level is the cloud computing level which provides the SaaS accessed by users using a light client via web browser.
- The security is given when a point-to-point connection is created between the node that wants to join to the network and an existing node through which the new node will be authenticated. This allows validating the node that joins the network.
- Messages are protected through asymmetric cryptography and the establishment of network session keys.

Spontaneous Networks: The best way of obtaining computing resources and processing our information

How it works?

- When a node starts a sport bets software, the node is accessing to the service shared by the remote node.
- Many different sport bets services can coexist in the same cloud network. Any request, computation or consultation is performed by using the aforementioned secure system.



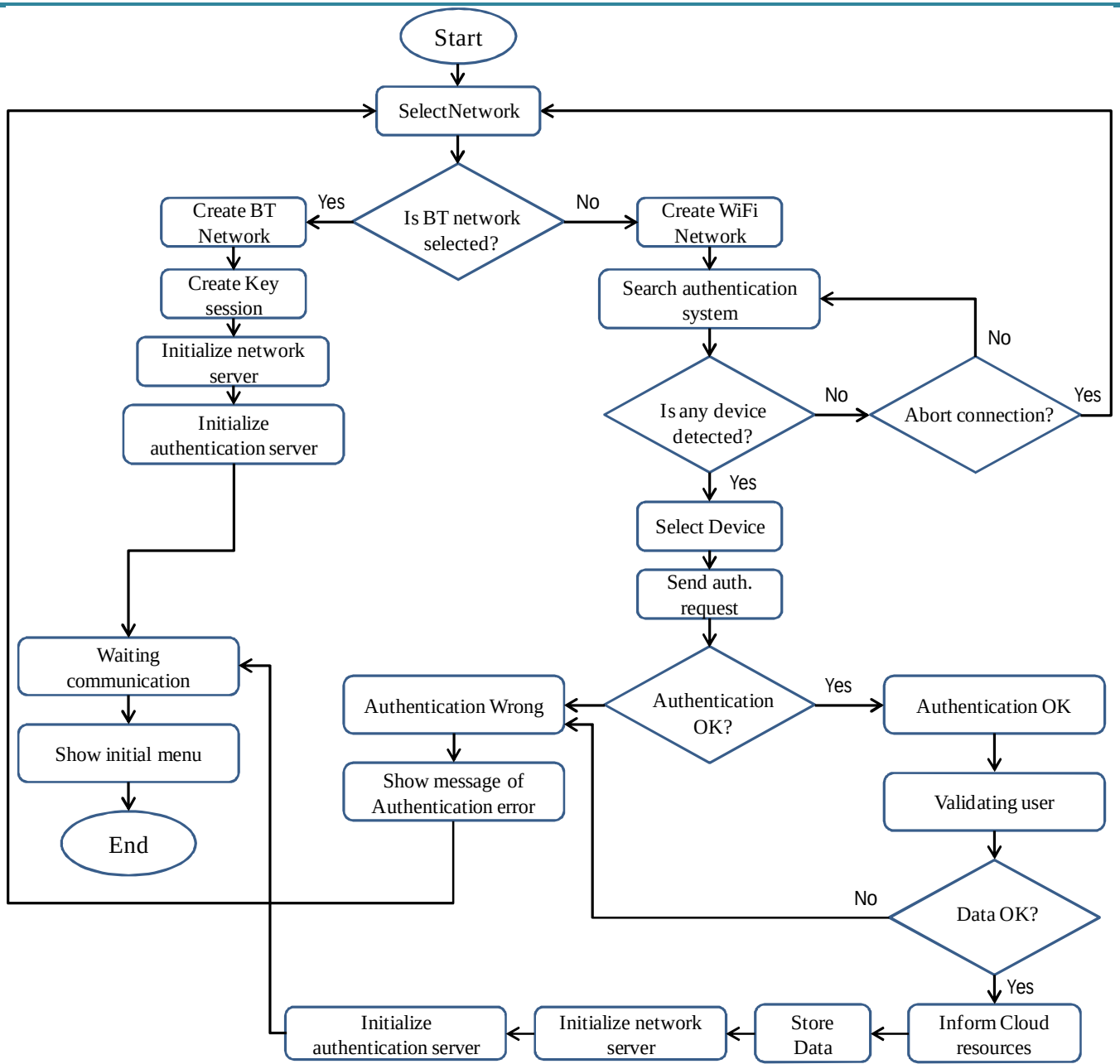


How it works?

- If the process is successfully performed, the system will store the data provided and the main menu will be displayed to select the network.

Spontaneous Networks: The best way of obtaining computing resources and processing our information

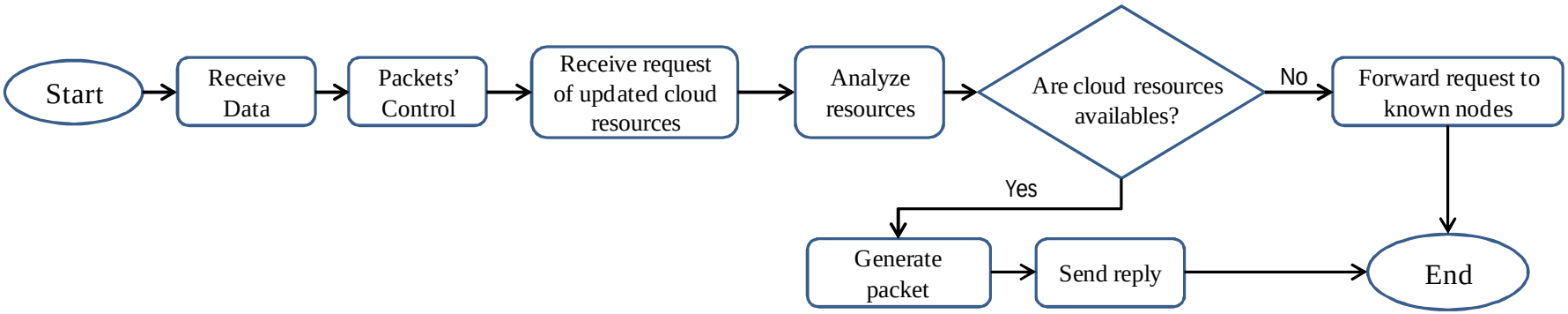
How it works?



Spontaneous Networks: The best way of obtaining computing resources and processing our information

How it works?

- In restrictive networks, the authentication process could be used to request information to the new nodes about their available resources or their installed sport bet software. In this case, the cloud network decides if this node is allowed to participate or not in the network (based on some predefined values).
- The acceptance of a new node depends on the node authentication and the data verification through a secure certificate and the data available at the network



Spontaneous Networks: The best way of obtaining computing resources and processing our information

ugr Universidad de Granada

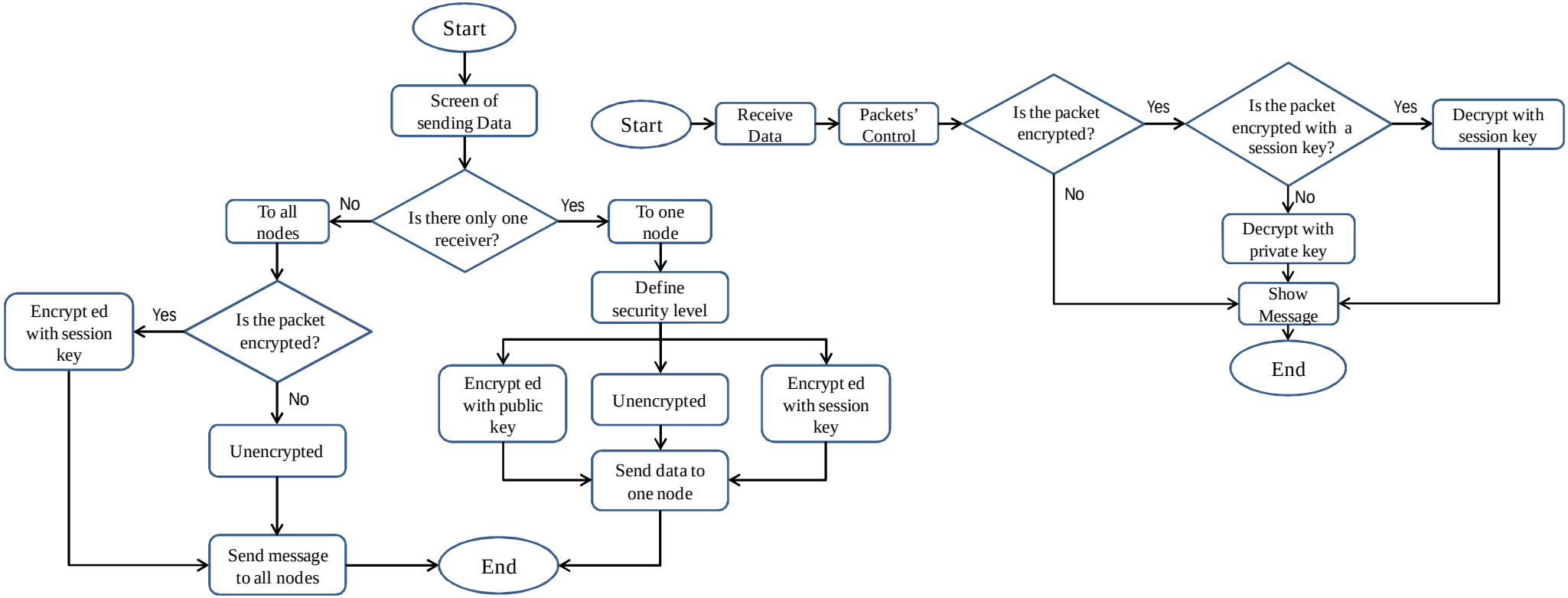
How it works?

- When a node is accepted, it informs to the mobile ad hoc cloud network about its resources and services including processing capacity and battery level.
- Node resources are shared in the network. When computing or storage are needed, any node can request resources to other nodes according to their availability of resources.
- Nodes can send update messages in order to know the current status of network resources. The control packet verifies the signature of the node that sends the information.
- In this way, the node will be authenticated. Updated information will travel signed and encrypted with the network session key.

Spontaneous Networks: The best way of obtaining computing resources and processing our information

How it works?

- When a node starts a sport bets software, the node is accessing to the service shared by the remote node.
- Many different sport bets services can coexist in the same cloud network. Any request, computation or consultation is performed by using the aforementioned secure system.



References

References:

[1] Raquel Lacuesta, Jaime Lloret, Sandra Sendra, and Lourdes Peñalver, Spontaneous Ad Hoc Mobile Cloud Computing Network, The Scientific World Journal, Vol. 2014 (2014), Article ID 232419, 19 pages. Available at: http://dx.doi.org/10.1155/2014/232419

[2] Sandra Sendra, Raquel Lacuesta, Jaime Lloret, Elsa Macias-López, A Secure Spontaneous Mobile Ad Hoc Cloud Computing Network, Journal of Internet Technology. Vol.-, Iss. -, No.-, **(In Press)**

Spontaneous Networks: The best way of obtaining computing resources and processing our information

Sandra Sendra Compte

Assistant Prof. At University of Granada (Spain)
(ssendra@ugr.es)