Informatics at the crossroads of Academia Europaea

H. Maurer
Graz University of Technology
Austria
Paris, September 21, 2011

hmaurer@iicm.edu
www.iicm.edu/maurer
http://www.ae-info.org/ae/User/Maurer_Hermann

Presented at: 23rd Annual Congress - “Chemistry, Sciences, Culture and Society in the making of Europe”
With the support of the Natural Sciences Section, UNESCO
Four main issues:

1. Informatics is being used more and more in all areas covered by Academia Europaea

2. Each area provides inputs for the development of entirely new techniques in Informatics.

3. Independent of such inputs new developments in Informatics are going to change our world still much more than most are aware of: „We have seen nothing yet.“

4. Typical areas are: WWW, and the wisdom and power of the crowd phenomenon (social networks). And two little known developments that will dominate Informatics theory and applications in this century: Information integration and pattern recognition in large, high dimensional data-sets.
John von Neumann (inventor of the concept of program stored computers) said in 1949:

„Most of what computers will ever be able to do we can do now, or will be able to do soon“

Cross judgement: Did not foresee power of networks, let alone WWW, did not foresee the rise of computer graphics, of new input and output devices, of large databases, etc.
Many scientists, inside and outside Academia Europae are telling me: „In a way, I really dislike computers: they are user unfriendly, maybe with the exception of some Apple products.

However, I have to use them and I am slowly getting more and more familiar with what they can do for me.“

Again, a cross misjudgement: new developments are coming at a speed so that it is almost impossible to keep up with them.

My favorite saying: *Whenever you have understood something new about computers it is bound to be obsolete.*

Let us now look at some of the sections in Academia Europea for a few examples of what I have said.
Section A1: Archeology and History

I am one of the reviewers for the ACM Journal on Computing and Cultural Heritage. A recent paper „Last House on the Hill…“ (Ashley, Tringham, Perlingieri,…) describing a major archeological excavation/discovery in Turkey says it all:

„Without proper software tools it would have been absolutely impossible to link all the data found by a large number of archeologists together in a holistic picture of the excavation, starting from the reconstruction of dozens of amphoras to finding the relationships between items and discovery of new aspects.“

See http://www.archaeovault.org/lhoth/
Using Google Earth a so-far huge unknown city „El Purgatorio Alto“ in Northern Peru covered by sand was discovered by Hasso Hohmann in 2009:

On the ground nothing obvious to see

Entrance to Castillo Chaukillo
Shard with puma-face

Row of 7 observatories

6,000 Years old!
A2: Classics and oriental studies
A3: Linguistic studies
A4: Literary and theatrical studies
A5: Musicology and history of art and architecture
A6: Philosophy, theology and religious studies
A7: Behavioural sciences
A8: Social sciences
A9: Economics Business and Management Sciences
A10: Law

Some members in each of the above sections are both heavily using computers (not just WWW) but also are involved in areas that have deeply influenced Informatics. Examples:

**Classical:** 3 D digitization of parts of object for correct measurements allows often reconstruction of whole object,…

**Linguistic:** Chomsky, language translation, ontologies,…

**Literary studies:** Textual comparisons,…

**Architecture:** New 3 D design tools like the CAVE and DAVE,…

**Theology:** Hypertext,…

**Social sciences:** Study of and using the Internet,…

**Economics:** Mathematical/computer models,…

**Law:** Databases,…
I guess no arguments necessary that those areas need Informatics and contribute to Informatics.

However, did you know that Turing’s proof of the undecidability of the „halting problem“ shows limits of what one can do with computers? It was already established in 1936! That strong encryption and data compression are only possible using deep mathematics?

Conversely, did you know that the famous mathematical four color problem has only been solvable using computers?

I am sure that most of you know that many constructs in physics, from general realtivity theory to cosmology to the duality of light can only be understood mathematically, but are considered beyond human intuition?
Access to information and information provisioning process have dramatically changed
- Not only catalog-based searches
- Portals enabling user-centered searches over heterogeneous document collections and topical databases
- Each domain requires different workflows
  - Chemistry: Entity centered
- Entities represented by
  - structures / images
  - string representations

B4: Chemistry Searching is always hard, but particular in Chemistry!

5-[2-ethoxy-5-(4-methylpiperazin-1-yl)sulfonylphenyl]-1-methyl-3-propyl-4H-pyrazolo[5,4-e]pyrimidin-7-one
Input for Search Engines difficult!

- Phenoxy methane
- Anisole
- Methoxybenzene
- Phenyl methyl ether
Problem is well known in the domain of chemistry
- Specialized structure based indexes established
  - Manually generated and expensive (big player: CAS)
  - Need for specialized search interfaces
  - Only experts can properly search for information

But still a Problem
- Not feasible for Open Access Journals and DBs
There are a plethora of file formats
- XML, HTML, PDF, Word …

Conversion of these formats to SciXML is usually not difficult
- PDF, however, is more complex
  - Every character has absolute position

Resulting in several problems
- Multiple line problem: 4-(aminomethyl) cyclohexamine
- Sub- and superscript: (1,7,7)-Trimethyl-tricyclo[2.2.1.0².⁶]heptan
- Fragments from tables and figures

For searching, complex conversions are necessary
More and more searches are done via Web Search Engines

Domain of chemistry particularly difficult:
  - manual maintained metadata, special search interfaces, expensive access to such databases

One goal: open up hidden chemical corpora by enabling text search via commonly used Web search interfaces

Slides based on research by Tonnies Sascha, L3S, Hannover, Germany

www.L3S.de/~tommies

This was just a tiny example that shows that every field (and chemistry is a good example) offers new challenges for Informatics
B5: Cosmic and Earth Sciences

We had a conference in Graz/Austria organized jointly by B5 and Informatics. The titles of many B5 talks say it all, all use Informatics methods heavily.

Evgenii Burov: Recent Advances in Numerical Modelling in Earth Science

Sierd Cloetingh: Challenges of Earth and Cosmic Sciences to Informatics

Taras Gerya: Innovative Solid Earth Modelling

Hans Suenkel: Gravity Field Determination from Space

Don Dingwell: Magma- The Ultimate Materials Modelling Challenge
All use Informatics heavily, and create stimuli for informatics. Let me just give one or two examples from medicine: first, the Liver Planner.

Example taken from Horst Bischof, Institute for Computer Graphics, Graz University of Technology.
Data-rich, individualised medicine poses unprecedented challenges for ICT, in hardware, software solutions.

We propose a data-driven, individualised medicine of the future, based on molecular/physiological/anatomical/environment data from individual patients.

We shall make general models of human pathways, tissues, diseases and ultimately of the human as a whole.

Individualised versions of the models, produced for each patient, will then be used to identify personalised prevention/therapy schedules and side effects of drugs.
The Future of Genomic Data

Factors of 10 since 2005 for genomics

Moore's law 1.5x/yr for electronics

- 2010 $1-7K (CGI) human genome
- 2004: $400M (ABI)
- 2000: $3 billion
Make out of Data Models
Flagship Goals

Years 1-5

Establishment of integrated molecular/anatomical prototype models of man, development of IT techniques to individualise these models based on high throughput data sources

Years 5-10

Development of infrastructure for model-based individualized medicine.

Interaction with relevant stakeholders/governments/healthcare and insurance systems to implement this approach throughout the healthcare system
24 Partners

- Max Plank Institut for Molecular Genetics
- Medical University Graz
- University College London
- Free University of Amsterdam
- University of Manchester
- Maastricht University
- EMBL
- Wellcome Trust Sanger Institute
- Kungliga Tekniska högskolan
- Imperial College London
- CIRMMP
- International Prevention Research Institute
- Uppsala University
- University of Luxembourg
- University of Leicester
- HARVARD Medical School
- University of Auckland
- Universite de Geneve
- Centro Nacional De Análisis Genómico
- Siemens
- Alacris Thoranostics GmbH
- Charite Universitätsmedizin Berlin
- Illumina
- Commissariat a l'énergie atomique et aux energies alternatives

(Slides with permission from Dr. Heimo Müller, Medical University Graz)
Still little known but very important areas, „Tasks for this century“

**Information Integration**
--- pulling information together from various sources
--- requires „sentiment“ analysis
--- remains Science Fiction unless new structuring paradigms are used in most areas
--- doable in some special cases (references of scientific papers, biographies)

**Pattern recognition in large multi-dimensional data-sets**
--- key for prognosis of disasters (in the physical world, in medicine,…)
--- key for non-invasive brain-computer interfaces
--- key for understanding long open questions

**Dark side in both issues.**
--- Ethics, cognitive psychology, social sciences and other areas have to play a more important control function than they do now
Of course all areas rely more and more on the Internet and the WWW.

Hence, also Academia Europaea needs the Internet for at least three areas:

1. Presentation of its aims and the aims and achievements of both sections and members for attracting support of organisations and new members

2. Nomination procedure has to be streamlined

3. Discussion forums open for all (?)

Item 1 is addressed by our server www.ae-info.org but it needs your help. I hope Wroclaw will play a big role in this respect

Item 2 has advanced quite a bit. What is still missing is automatic update of www.ae-info.org based on accepted nominations.

Item 3 is still wide open: Should the general public be allowed? Should this also be a tool for publication? Who will be able to define which „closed user groups?“, etc.

Let me finish to point you once more to our www.ae-info.com server
Thanks for your attention. Direct any queries to hmaurer@iicm.edu