

Discussion of “Biomedical Informatics: We Are What We Publish”

A. Geissbuhler¹; W. E. Hammond²; A. Hasman³; R. Hussein⁴; R. Koppel⁵;
C. A. Kulikowski⁶; V. Maojo⁷; F. Martin-Sanchez⁸; P. W. Moorman⁹;
L. A. Moura¹⁰; F. G. B. de Quirós¹¹; M. J. Schuemie¹²; B. Smith¹³; J. Talmon¹⁴

¹Department of Radiology and Medical Informatics, Geneva University, Geneva, Switzerland;

²Duke Center for Health Informatics, Durham, North Carolina, USA;

³Department of Medical Informatics, AMC-University of Amsterdam, Amsterdam, The Netherlands;

⁴The Biomedical Informatics Center of Excellence, Information Technology Institute, Ministry of Communications and Information Technology, Egypt;

⁵Sociology Department and the School of Medicine, University of Pennsylvania, Philadelphia, USA;

⁶Department of Computer Science, Rutgers – The State University of New Jersey, New Jersey, USA;

⁷Departamento de Inteligencia Artificial, Facultad de Informática, Universidad Politécnica de Madrid, Madrid, Spain;

⁸Health and Biomedical Informatics Centre, The University of Melbourne, Melbourne, Victoria, Australia;

⁹Medical Informatics Department, Erasmus University Medical Center, Rotterdam, The Netherlands;

¹⁰Assis Moura eHealth, Porto Alegre, Rio Grande do Sul, Brazil;

¹¹Department of Health Informatics, Hospital Italiano de Buenos Aires, Buenos Aires, Argentina;

¹²Janssen Research & Development, Titusville, New Jersey, USA;

¹³Department of Philosophy, University at Buffalo, Buffalo, New York, USA;

¹⁴Centre for Research Innovation, Support and Policy, Maastricht University, Maastricht, The Netherlands

Keywords

Biomedical informatics, natural language processing, Medline, ontology

Summary

This article is part of a For-Discussion-Section of *Methods of Information in Medicine* about the paper “Biomedical Informatics: We Are What We Publish”, written by Peter L. Elkin, Steven H. Brown, and Graham Wright. It is introduced by an editorial. This article contains the combined commentaries invited to independently comment on the Elkin et al. paper. In subsequent issues the discussion can continue through letters to the editor.

Correspondence to:

See list of authors' addresses at the end of the article.
Methods Inf Med 2013; 52: 547–562

With these comments on the paper “Biomedical Informatics: We Are What We Publish”, written by Peter L. Elkin, Steven H. Brown, and Graham Wright [1], the journal seeks to stimulate a broad discussion of biomedical informatics as a discipline. An international group of experts has been invited by the editor of *Methods* to comment on the paper. Each of the invited

commentaries forms one section of this paper.

1. Comment by A. Geissbuhler

One of the striking characteristics of our field of biomedical informatics is its recurring attempt at defining itself, too conscious of its polymorphic and rapidly evolving nature, and somewhat afraid to being just considered as an interface between domains rather than a true discipline based on an original body of knowledge and skills.

It is therefore not surprising that experts from our discipline, which sometimes is defined as being mainly preoccupied with the meaning of information and its proper computability and use, should exploit the precise tools that were developed to manage information, meaning, and knowledge to carry out the introspective task of defining its essence.

These very tools, ontologies, natural language processing systems, and other concept extractors have all been studied in detail by biomedical informatics specialists. Although they have been refined over decades, they still lack widespread usage within health information systems. We know

what can be expected from them, we also know their limitations, and for example how difficult it is to represent common sense or extract actual knowledge from information without a significant human intervention. This “demise of the Greek Oracle model”, coined by Miller [2] about medical diagnostic systems, probably applies to most domains of biomedical information classification and analysis, when attempting to emulate intelligent behaviour.

So, just like the best results are obtained when creating synergies between knowledge bases and domain experts, the methods described in this article could be used to shed some light on some other, more specific issues.

For example, if topics of interest described in the Health Informatics Ontology do not seem to be covered by the typical literature of our field, where are they actually addressed and published? We know that many significant publications involving biomedical informatics actually get published in an ever wider range of medical journals. This is good news, provided that we can track them.

We also know how weak we are at demonstrating the impact of our interventions, in a way that would be convincing outside of our own community. So, could we, with

the tools described in the article, identify important publications, particularly those demonstrating impact on health outcomes, where our field would not easily be identified as the prime contributor, but would play the role of a significant enabler?

These would be important answers to difficult questions, and yet another way of affirming the relevance of our discipline.

2. Comment by W. E. Hammond

“I don’t know what you mean by ‘glory’”, Alice said.

Humpty Dumpty smiled contemptuously. “Of course you don’t – till I tell you. I meant ‘there’s a nice knock-down argument for you!’”

“But ‘glory’ doesn’t mean ‘a nice knock-down argument’” Alice objected.

“When I use a word”, Humpty Dumpty said, in rather a scornful tone, “it means just what I choose it to mean – neither more nor less.”

“The question is”, said Alice, “whether you can make words mean so many different things.”

“The question is”, said Humpty Dumpty, “which is to be master – that’s all.”

Alice was too much puzzled to say anything, so after a minute Humpty Dumpty began again. “They’ve a temper, some of them – particularly verbs, they’re the proudest – adjectives you can do anything with, but not verbs – however, I can manage the whole lot! Impenetrability! That’s what I say!”

In Lewis Carroll’s “Through the Looking-Glass” [3] Humpty Dumpty discusses semantics and pragmatics with Alice.

We, in informatics, seem to have a difficult time defining the words associated with what we do. Examples include the core of our existence – informatics – as well as such terms as what most of us now call the electronic health record. This paper provides an interesting approach to this matter. We also seem to have a tendency to add an adjective or letter to what word we use to define what we are doing. Examples include e-health plus a series of nouns following “e-”; m-health plus a series of nouns following “m-”; and i-health plus a series of

nouns following “i-”. Founding members of the American College of Medical Informatics (ACMI) spent four hours discussing (I might say arguing) what to call this evolving field. The word informatics derived originally for the Russian term *informatik* and the French *informatica*. The Germans had also begun using that term. Medical seemed to be a reasonable addition, and the field was defined as “Medical Informatics”. Unfortunately, the Nursing Community strongly associated the term “medical” with physicians, and the community began using the term “clinical informatics”. My current count is that 20 adjectives have been added to the word informatics. Even more interesting is the emergence of a new term – Data Science. I like this word because it seems to avoid the domain adjective that precedes informatics.

I congratulate the authors in addressing a topic of some importance. It is an interesting approach to defining the field. Unfortunately, I fail to understand how to use the findings of this paper. If it is an attempt to define the field, how does that relate to the definitions of the experts in the field. First, the results seem likely to differ, depending on the source of the terms for comparison. Use another controlled terminology other than SNOMED-CT, the results are likely to be different. Using MeSH should also yield different terms, although MeSH is derived directly from the literature itself. MeSH terms continue to grow as the field evolves. In fact, MeSH in itself would likely yield equivalent results to the natural language approach. The study also does not take into consideration the evolution of technology, applications, domain, and research in informatics. It would be interesting to use this method year by year to measure the evolution of the field. In the late 1980s and early 1990s, attempts were made to bring the clinical informatics and the molecular biologists together. The meeting was most interesting, but there was no follow-up. One of the AMIA Annual Symposia had a molecular biology theme. Again, the idea did not catch on. ACMI inducted a number of molecular biologists into ACMI but again will little mating. Finally, in recent years, the Joint Summit on Translational Science (TBI/

CRI) appears to be having success in bringing the groups together. One would assume papers with terms representing both domains would begin to appear together. Doing this analysis on a year by year basis should help define the growth of the field.

One problem with this approach is that the terms being used for comparison have been defined for different purposes. Clearly SNOMED CT was not created to define the field. Ever taking the words for the literature will not reflect the broadness of the field.

The basic questions I wished the authors had addressed would be the purpose of the definition of the field. Is the purpose to explain to others the domain of our work? Is it to influence the curriculum that is taught? Is it to influence the kinds of jobs our graduates get? Or is it to now meet the requirements for certification.

I appreciate the opportunity of commenting on this article. My belief is that it will stimulate much conversation among our colleagues. But I end as I started. **“When I use a word”, Humpty Dumpty said, in rather a scornful tone, “it means just what I choose it to mean – neither more nor less.”**

3. Comment by A. Hasman Introduction

In their paper “Biomedical Informatics: We Are What We Publish”, Peter Elkin and colleagues evaluate a top-down approach to define biomedical informatics [1]. Both IMIA and AMIA produced a consensus document defining the field. The authors investigated the question whether the consensus documents contain the necessary concepts to index medical informatics papers. If this is not the case, they propose that a bottom-up approach in which concepts from the medical informatics literature will be mined, will lead to a better description of the field.

In order to investigate the above mentioned question, from the IMIA and AMIA consensus documents concepts and terms were derived, analyzed and merged, resulting in a Health Informatics Ontology (HIO). The HIO was then built into a terminology server of an intelligent natural

language processing system and used to index medical informatics articles obtained via Pubmed.

Discussion of the Paper

I have a number of methodological questions concerning the study. The questions concern the adequateness of both the HIO and the abstracts. The authors themselves already indicated some limitations of their paper.

The Adequateness of the HIO

With respect to the HIO one can ask the question whether the IMIA and AMIA documents lead to an optimal ontology describing the current insights about the topics covered by the field. The documents contain rather abstract terms and that probably will lead to a HIO also containing terms that are partly too abstract to suitably index the medical informatics papers.

Another problem may be the fact that the HIO does not seem to contain many synonyms. There are 433 concepts and 462 terms, so there are almost as many concepts as there are terms. Especially because the underlying hypothesis states that if not all papers can be indexed, this means that the HIO is not complete and that therefore the field is not covered by the HIO, the lack of synonyms may be particularly serious: could the incompleteness not be due to this lack of synonyms and not so much be due to missing concepts in the HIO?

The authors should have checked this. They could have selected a sample of the abstracts that could not be indexed by the HIO and then check the reason why the system could not index them: were less abstract terms used in the abstracts, were synonyms used or were terms used in the abstract really missing in the HIO.

The Adequateness of the Abstracts

Another question one can ask is whether the medical informatics papers obtained from Pubmed are representative for the field. If not, this does not so much influence the results presented in the current paper but it is important when one wants to use the titles and abstracts of Pubmed

for defining the field. Medical informatics is a field that also comprises for example the fields of systems analysis and software design and also software development methodologies are involved. Are these domains adequately covered by Pubmed? I do not think so.

Another question is what is described in the abstracts. Were all abstracts structured and if not, which percentage of the abstracts was unstructured? Unstructured abstracts perhaps may say more about the scientific results obtained in the study and therefore contain more terminology from the medical field than from the medical informatics field. In the 27,000 abstracts that were indexed by SNOMED on average more than one term was found in each abstract. However, the authors do not mention which percentage of the titles and abstracts did not contain SNOMED terms. In my opinion structured abstracts will provide more information about the methodology used and therefore refer to more medical informatics aspects than unstructured abstracts.

And again, the terminology used in abstracts and titles may be less abstract than the terms used in HIO or contain synonyms not present in the HIO.

The Author's Conclusions

The authors determined that with HIO only 37% of the abstracts and titles could be indexed. The fact that only 37% of the papers contained terms from the combined AMIA/IMIA document does not mean, in my opinion, that the other 63% of the abstracts and titles contain other concepts not yet considered in the combined ontology. As said above, the terminology of the HIO may not be complete, lacking not so much concepts but terms and synonyms at an appropriate level of abstraction. Also it may be the case that the IMIA and AMIA documents are not totally suitable for the purpose. I wonder why the IMIA Recommendations on Health and Medical Informatics education were not also used to obtain the HIO. Of course the critique that these documents are not totally suitable may also support the authors' suggestion to approach the definition of the field in a bottom-up way.

Conclusion

In my opinion the idea of indexing biomedical informatics papers with terms obtained from consensus documents may indeed provide insight in whether the top-down approach will lead to a complete HIO or that in addition a bottom-up approach is needed. The results in the article do not yet settle the question. The appropriateness of the HIO should be investigated. Also if the HIO obtained by the top-down approach is not complete and a bottom-up approach is needed one should not rely on Pubmed alone. Also other literature databases in fields related to biomedical informatics should be considered.

A bottom-up approach in itself probably will lead to a more detailed ontology of our field than a top-down approach. Such a bottom-up approach should be further developed.

Especially for educational purposes having a detailed ontology of the field will provide us with the opportunity to evaluate educational programs and to determine what parts of the ontology are covered by a program. The availability of such an ontology would for example be a useful feature in the accreditation process of health informatics programs.

4. Comment by R. Hussein

The paper published recently in *Methods* entitled "Biomedical Informatics: We Are What We Publish" [1] analyzes the published literature in Biomedical Informatics (BMI) in order to provide a better understanding of the breadth and depth of the BMI field. The performed analysis provides a clear insight about how the different concepts of Health Informatics Ontology (HIO) are addressed in the BMI literature. The results of the analysis show that only 37% of the used literature in this study has at least one concept of the HIO concepts. This was also followed by a list of the most frequent HIO concepts identified in the used literature. Apparently, the technical aspects, such as standards, technology, systems evaluation, and networking, etc. come on the top of the list of the most used HIO concepts. As expected, the human-based

concepts, particularly the user interface, appear in the last places of the list.

Besides literature, international organizations recently led several initiatives addressing the challenges and achievements in transforming the health informatics research into the eHealth domain. The most significant initiatives in this regard:

- eHealth-conference series (2007–2013): under the theme of "Health Informatics meets in eHealth – From Science to Application and back" [4]. During this series, the researchers discussed the health informatics methods and their applications in eHealth.
- Making the eHealth Connection conference series (2008): under the theme of "Global Partnerships, Local Solutions" [5]. This initiative was led by the Rockefeller Foundation in coordination with American Medical Informatics Association (AMIA), Health Level Seven (HL7), International Medical Informatics Association (IMIA), World Health Organization (WHO), and others. The conference aimed at advancing eHealth and improving health systems in the developing world.

In addition, WHO conducted a global survey in 2009 to address the state of eHealth development in the 114 WHO member states. The eHealth indicators, including the eHealth foundation actions as well as eHealth applications on national level, are presented in the form of country profiles [6]. In 2011, WHO and International Telecommunication Union (ITU) published "National eHealth Strategy Toolkit" to facilitate the development of an effective national eHealth policy and strategy [7].

At the academic level, IMIA published the revised version of its recommendations on Biomedical and Health Informatics (BMHI) in 2010 [8]. The recommendations specify the required educational needs, in terms of knowledge and skills, for healthcare professionals in utilizing Information and Communications Technology (ICT) in medicine and healthcare. Furthermore, the AMIA published its whitepaper in 2012 that describes how the Biomedical Informatics (in the basic research domain) synergistically and interchangeably interacts with Bioinformatics, Imaging

Informatics, Health Informatics and Translational Informatics (in the applied research and practice domain) [9].

To build on the outcomes of these conferences and initiatives, the authors are recommended to also analyze the HIO concepts in the eHealth domain in particular. This means, applying the same methodology by downloading the eHealth literature and then parsing the downloaded literature with HIO concepts. Consequently, the results can be utilized as follows:

1. Creating the list of the most frequent HIO Concepts identified in the eHealth literature. This analysis will give a clear insight on to which extent BMI facilitate the adoption of eHealth.
2. Categorizing the created list of HIO concepts in eHealth by country/region, where the eHealth study was conducted. In this way, a complementary view to the WHO Atlas (eHealth country profiles) will be provided.
3. Indexing the literature of BMI education and capacity building. This literature could be reviewed in light of the IMIA recommendations on BMHI education as well as AMIA core competencies for graduate education.

Lastly, analyzing the HIO concepts in the eHealth domain has threefold advantages:

1. Providing a clear insight of the current eHealth status in each country. This will encourage the adoption of the eHealth toolkit, particularly in the developing world where customization of eHealth solutions is necessary.
2. Developing the country profiles of the BMHI competencies. This will facilitate North-South and South-South collaboration, sharing expertise between the different countries, and partnering in regional/international projects.
3. Reviewing the existing BMHI capacity building programs from both educational and professional perspectives. This will lead to refining the IMIA recommendations on BMHI education and AMIA core competencies for graduate education according to the real-world needs and future trends in eHealth.

5. Comment by R. Koppel

This article by Elkin, Brown, and Wright [1] enhances our understanding of our discipline. It could have been alternatively entitled: "Do we know what we are talking about when we talk about biomedical informatics?" Another possibility, referencing the article's findings on the conceptual-linguistic mapping in the literature, might be: "Research reveals almost 40% consanguinity of terms and literature in the medical and biomedical informatics oeuvre: a glass half full?"

Wordplay aside – an odd term given the work of the article – the first reward of this important work is in clarifying the scope, focus, central tendencies, and richness of the medical and biomedical informatics research literature. This would be enough, helping us better see what our discipline is, where it is deep, where it is clustered, where it is thin, where there may be gaps. But the article also gives us a measure of the diversity and the sometimes fragmented aspects of the discipline; a measure of the *reciprocal* of focus, where a large accumulation of loose threads still hang from the core fabric of the discipline. This scattering of less fully absorbed topics and literature offers three possibilities:

1. We can cherish the opportunity for including some of these prodigal subjects more fully into the corpus of the discipline. The large number of works and concepts that would benefit by greater and more active inclusion achieved through a broader perspective of medical and biomedical informatics. Such a strategy also provides us opportunities to expand and encompasses a wider purview of medical informatics. More, the mutual learning achievable would be of value to both core and "peripheral" areas.

And/Or

2. We can complain that we have been lousy disciplinary imperialists, allowing a disciplinary free-for-all with many "stray" ideas and work that should be more closely and logically entwined into our grasp. Following this logic, we have failed to claim that which is rightfully ours; not because we are unambitious,

but because our work includes so much of medicine, biology, computer science, engineering, administration, management, human factors, training, marketing, regulation, etc. There's just so much to pull together that we are spread thin and fail to benefit as much as we should by the broader purview that needs a systematic gathering into the fold. This work by Elkin, Brown and Wright is a valuable step in recognizing and perhaps starting to encompass that which is in our discipline's orbit but not sufficiently in our close gravitational pull.

And/Or

3. We can take the findings from this research and simply acknowledge that medical- and bioinformatics is all over the place: diverse, wide-ranging, and rich, but just too darn hard to cluster into a neat rubric.

I, for one, reject the third option, and argue that as messy as the discipline appears to be, there's more that is centripetal than centrifugal in medical and bioinformatics. Consider biology: it's as broad and expansive as our discipline, but no one is suggesting it's too difficult to encompass its concepts and literature in one whole.

What's Missing, What's Underserved?

The research also highlights some aspects of medical- and bioinformatics that deserve more attention. For example, one of the great failures of the entire healthcare IT effort is the lack of interoperability. Yet the term receives little attention in the listing of concepts and literature. On the other hand, two other major disappointment of healthcare IT, usability and data standards, are amply represented, at least in the literature, if not in reality.

In sum: This work, "Biomedical Informatics: We Are What We Publish", helps us understand ourselves, and our linkages to others in our field and in related fields. It's an essential contribution to our knowing what we are doing and what we are trying to do. It also helps us know where to look for useful concepts and literature perhaps not directly in our sights, but nearby.

The authors are to be thanked for helping us see who we are and where we are situated in the universe of ideas and research.

6. Comment by C. Kulikowski

This paper [1] describes how the authors arrive at a possible consensus definition of Biomedical Informatics through an analysis of papers gleaned from a literature search using the terms Medical Informatics and Bioinformatics, which yielded 153,580 and 20,573 articles respectively, with 5,855 overlapping articles including both terms. They report that 37% of these articles had titles and abstracts containing at least one concept from a Health Informatics Ontology (HIO), and indexed the articles by SNOMED CT terms to determine their medical or clinical topic, as well as by the Gene Ontology (HGNC) to correlate with gene-related factors referenced in the bioinformatics literature. The authors claim that their analysis adds value to two expert consensus driven definitions of Biomedical Informatics that have been developed by AMIA and IMIA for their core competency and educational program description documents.

There are several critiques that can be made of this article, which lacks clarity and specificity in its study design, especially in connection to the scope and assumptions made in the selection of terms from the different reference sources and in the methods used for the analysis. Based on the evidence and arguments presented, its thought-provoking title would not appear to be justified.

A first critique is that the authors have not defined exactly what is the problem they are trying to solve nor the questions they are seeking to answer with their analysis – other than showing that a "bottom-up" extraction of terms from the literature should help complement the "top-down" organization of the discipline of Biomedical Informatics by the professional societies and standards groups. This is in itself not controversial, and could indeed yield useful insights, but the way in which the authors describe their study raises many questions of methodology and specific design used,

so as to make their results confusing and unclear at best, and of dubious use at worst. Their reliance on titles and abstracts alone is a major limiting factor, as is their use of terms taken from sources designed for very different purposes.

The initial criteria for the selection of terms used by the authors for their search are stated to have been only "Medical Informatics" and "Bioinformatics", which leaves out many relevant articles that might have been found using the terms "Health Informatics", "Nursing Informatics", and makes one wonder why they did this. A hint comes from their paper where they describe how bioinformatics was not included in the IMIA Knowledge Base of one of the co-authors (Wright), so it might seem that this study is intended to show the additional wealth of terms that could be added if the bioinformatics literature were searched also. However, the study design is not nearly sufficiently specified to make it possible to say whether this would be the case or not, assuming that the authors did have this in mind.

A central critique is that the paper does not describe how the Terminology Merge Utility (TUM), developed to assess the terminological coverage of SNOMED-CT for clinical problem lists was actually used in the present study, and, most importantly, what might be some of the assumptions made explicitly or implicitly in merging with the very disparate and differently-motivated AMIA and IMIA terminologies. As result, the merged Health Informatics Ontology (HIO) that the authors used to interpret the results from their parsing of the literature is difficult to assess in terms of completeness of coverage or specificity/sensitivity for retrieval of the kinds of lower-level concepts from articles as they next proceed to do. And, the decision to use a random subsample of 27,000 articles from the already undefined and potentially-inadequate initial sample of 168,298 titles and 121,561 abstracts (why the difference?) parsed with the HIO using the iNLP terminology server is only justified on grounds of conserving server time (3 days, which while understandable in practical terms, raises questions about the results in relation to the potential bias of what was or was not included, how representative or

random was the sample, and how were these or other properties of the sample determined?). The picture in Figure 7, presented as an “example of a conclusion” does not make clear what it is intended to explain, and the distinction between concepts and nodes listed in Table 1 is not described, and neither are their implications for differences found between IMIA and AMIA terminologies and the merged HIO.

The paper also states that: “*The Bioinformatics articles were also parsed with the Human Gene Nomenclature Committee Ontology (HGNC Ontology) to determine the breadth of human gene coverage attributable to the Informatics literature and how well HGNC Ontology represented the content of bioinformatics articles.*”

Descriptive statistics were applied to the resulting dataset of codified Informatics literature.”

However, no more is said about how this was specifically done, and the only reference in the conclusions is:

“Of the 20,573 Bioinformatics articles, 14,427 had HGNC Ontology codes. Of the 26,953 human genes only 3,275 were identified in this corpus of Bioinformatics literature. When we combine the HIO and HGNC Ontology we find coverage of 76,671 (45.6%) of the 168,298 unique articles.”

The significance of the above conclusion is unclear and not relatable directly to the long lists of concepts from the IMIA and AMIA terms and their merged IHO, which are given as samples only, and related from the study of the subsample of 27,000 articles to inclusion of body-part terms from SNOMED-CT codes, for which the authors state:

“To evaluate the overlap of Biomedical Informatics with clinical medical specialties we parsed 27,000 articles with abstracts and the most frequent areas by body site identified are listed in Table 3. Overall there were 37,141 occurrences of SNOMED CT clinical concepts in these 27,000 titles and abstracts. Body site areas without representation in the Biomedical Informatics literature are listed in Table 4”

and because “only 37% of articles titles and abstracts were identified using the HIO”

the limitations admitted by the authors are not only pertinent, but also suggest that this is a very early and preliminary study

from which it would be premature to draw specific conclusions.

It would have been useful if the authors had done what they say is needed as future research in order for this paper to go beyond hinting at what are the promises of the type of analysis they have carried out. Not only would correlation with MeSH terms be helpful, but so also would a comparison with terms from expert reviews, as well as better ways of understanding the relationships between those in the published literature and the broader results contained in detailed technical reports at one extreme, and critical discussions of the literature at the other. The very definition of topics and subtopics varies by specialty and subspecialty, and the implicit goals and informal languages of the different specialty professional groups strongly affects their semantics, making the kind of superficial analysis presented in the paper by Elkin et al. difficult to interpret or use. Among the challenges raised by the author’s approach one could include the need for practical computational ontologies to become operationally updatable which is not addressed by this preliminary study, and is even barely within the state-of-the-art. The heterogeneity of available formal and informal models for both knowledge representation and reasoning and the frequently divergent objectives of scientific inquiry vs. those involving practical knowledge management in order to achieve “satisficing” healthcare objectives, present a formidable set of epistemological challenges that do not lend themselves to facile application of the currently immature “big data” approaches to the analysis of text and other media.

But, to conclude on a positive note, the present paper can hopefully serve the practical purpose of stimulating further debate during the long and likely never-ending journey of defining what our exceptionally broad and heterogeneous field of health and biomedical informatics can seek to become.

7. Comment by V. Maojo

“Biomedical Informatics: We Are What We Publish” [1] presents, right at the start, with its title, a thought-provoking and intri-

guing statement. Many philosophers, journalists, scientists, novelists, non-fiction writers, publishers, critics, etc. could write very different papers – and, probably very many alternative versions of such papers – dedicated to this idea based on their experiences in their own fields, bringing very different perspectives to this complex and basic question implicit in all research, scholarship, and art. In their manuscript, Elkin and colleagues present a challenging thesis: we, (bio)medical informaticians, can define our area based on what we publish. Such a strong statement seems to suggest one more attempt to answer a question that would have intrigued even Freud, given the great number of times that biomedical informaticians have asked themselves, for several decades, the underlying question: what is biomedical informatics (BMI)? We – all the people that work in the field and write and read medical informatics books, journals and papers – usually clearly recognize ourselves as biomedical informaticians, but this identity has been tough to define and explain to others, by specifying precisely, what it is that we really do. Are we biomedical or computer scientists, biomedical or pure engineers, information specialists or some kind of brokers between doctors and computer scientists or IT practitioners more generally? [10]. Is what we do primarily based on science, engineering, art or a mixture of all of them? [11].

To provide their answer to this recurring question, Elkin and colleagues gathered two current medical terminologies, developed and reported by groups of scholars belonging to the International Medical Informatics Association (IMIA) [12] and the American Medical Informatics Association (AMIA) [9]. Then, they merged both terminologies, creating a new integrated one, which became the basis for a mixture biomedical/health informatics ontology (HIO). Using this ontology, they worked with the results of two searches of the medical literature, performed with the terms “medical informatics” or “bioinformatics” in Pubmed. A random sample of 27,000 articles with abstracts from these results was parsed with the Systematized Nomenclature of Medicine - Clinical Terms (SNOMED-CT) to determine the breadth of clinical coverage attributable to the Bio-

medical and Health Informatics literature. The bioinformatics set of papers was parsed with the Human Gene Nomenclature Committee Ontology (HGNC Ontology) to determine the breadth of human gene coverage attributable to the Informatics literature. Finally, the authors propose how this work can facilitate further research in the field.

Leaving aside the value that the intrinsic focus of the paper has from reporting on a somewhat new ontology of the field, I find myself having to disagree with various statements made in the paper, which I find not necessarily true. The first one is implicit, in the sense that we can define one scientific field based on an ontological/taxonomical representation of the different concepts/categories/properties/semantic relations (etc.) that appear in such domain – an issue to which I will comment on below. Another statement, this one explicit, is that the SNOMED CT indexing used to develop the authors' terminology demonstrates a maturing of the BMI field as a health scientific discipline. Likewise, one can question the authors' consideration of the IMIA and AMIA classifications as authoritative in capturing the scope and depth of the field from the perspective of the organizations sponsoring the classification work. First, coincidences with SNOMED-CT only show semantic relations, not really scientific maturity. Maybe the strongest semantic relations could be made, ironically, with those aspects of clinical medicine that are less scientifically mature. Some deeper analysis should be necessary for such a statement. Second, the IMIA and AMIA classifications are just representative of a small sample of a community in the form of the committee that drafted the classifications, which, while maybe very well-informed based on the expertise of their members, is hardly the final word on how concepts, relationships and the structure of knowledge can be described in the field. The problem with all such professional society-sponsored activities, is that anytime that we label something as "authoritative" in science, we are almost inviting the brightest young minds in the disciplines involved to find exceptions and contradictions in parts of the ontology, ideally refuting important parts and

coming up with new theories to support a discipline. As result, a taxonomy used in practical scientific and technological fields cannot be falsified – it only aims to be a working, or operational representation of part of the important knowledge in the field.

To analyze the biomedical literature by searching Pubmed, aiming to extract useful information and use it to define a field like biomedical informatics is scientifically problematic. Papers are currently indexed in Pubmed manually, using concrete keywords assigned by the US National Library of Medicine staff. Such indexing tasks have been carried out for decades, involving many experts, almost surely each bringing their own personal, subjective expertise, criteria and variants of indexing methods to the task. Furthermore, the analysis and indexing of Pubmed papers has followed a different path over recent years, leading to new computerized methods of extracting and analysing the available information. The use of natural language processing (NLP) and text mining techniques has transformed how information can be retrieved from bibliographic databases [13, 14]. It is currently possible to analyze, (semi)automatically, all Pubmed abstracts – and many full papers –, improving on the simpler, traditional Pubmed searches. However, for this author, Pubmed searches – even if they were much more comprehensive than the work presented here – cannot be sufficient to carry out the necessary analysis for defining the BMI field and lead by themselves to new, original directions or insight. In addition, the reported good results of the NLP tools developed by the authors were obtained with other experiments and datasets. In fact, the somewhat poor results of some of the analyses carried out show the problems of the research design of this experiment. In addition, for these kinds of NLP and text mining experiments, it is still necessary to systematically create a corpus of papers, annotated by experts, which can serve as the basis for deeper analyses [13].

The main value of these two AMIA and IMIA classifications is to provide some consensus and a solid reference for further work and academic use. Merging both classifications can provide some additional in-

formation, surely, but the consistency of the new classification may be problematic, since the objectives, methods and approaches of the two previous taxonomies were surely different, and their results were assessed by groups of experienced BMI scholars and professionals differently and for different purposes. A direct merging of both classifications does not necessarily mean that the resulting classification will be better. Then, a thorough evaluation is necessary. For example, some examples obtained by the authors seem quite unspecific, and it is difficult to consider them as really significant for BMI. For instance, in the manuscript, some of the most frequent HIO concepts identified in the informatics literature do not seem to provide equally meaningful or representative information. E.g. "technology", "evaluate IS/IT", "networking", "evaluation", "guidelines", "neural networks", refer to terms which belong to different conceptual categories, with possible overlaps and are very generic, so they could be assigned to many different domains. All these classifications need some manual evaluation and further curation, and their relevance is, basically, statistical. To go beyond such limits, aiming to extract or infer new knowledge is a more difficult task. As for the analysis of what we call now "Big Data", we do not really need more and more data – or terms, or papers – but better hypotheses to analyse the available information with new perspectives, which may lead to new scientific discoveries or insights. If not, the fundamental, scientific problems, still remain there, probably hidden beyond any semantic analysis. However, the analysis of relationships between a comprehensive health informatics ontology and clinical research/care, can surely have specific applications of great practical use – e.g., annotation of research experiments or clinical cases, disease classifications, etc.

Conclusion (a Digression on Semantics)

Medical informaticians have usually and strongly rejected the superficial perception that many people from outside the discipline may have of medical informaticians – "you just apply computers to medicine,

don't you?”. This implies a direct, simple – or simplistic – explanation of what we do. It also implies an underlying view of the field as having only a technological or engineering focus, usually around software development. Such a straightforward definition avoids the deep scientific issues that we have been endeavoring to tackle in the discipline and would like to have informed people associate with (bio)medical informatics. But, haven't we avoided some of the deeper scientific problems that underlie medicine and information in at their provocative intersections? As stated in an article with my colleague Kulikowski elsewhere [15], we urgently need better theories to define the field of BMI from a scientific perspective. With any new central theory, the meaning of the underlying concepts and associated ontology may change. Then, any semantic effort must be acknowledged, but its temporary character must be recognized also. The semantic analysis of an entire field might show, for instance, how many people dedicate themselves at a given time to the same topics, what are the most popular keywords associated with their work, or how agencies have funded certain concrete subfields and not others. However, to define what is BMI is a task that still needs the contribution of people with a broader experience – often newcomers who are not afraid to disrupt the quiet life of professionals working in a domain, by proposing what we might recognize, years later, as “paradigm shifts” [15].

Many papers have been dedicated to the analysis of “what is BMI?” [16–18]. For this author, they all finally suggest that, from a scientific perspective, clear answers remain still very operational and contingent on developments in a rapidly advancing field at the boundaries of science, technology, and the art of its application to the deepest of human concerns – our health. In addition, Elkin et al.'s paper mixes the search for a scientific definition and characteristics of biomedical informatics with a description of the field from an educational, academic, professional or pragmatic “list of competencies” view. These differentiated aspects should be seen, too, from different perspectives. Such broad and deep analysis cannot be only based on the results of

Pubmed searches and subsequent semantics analyses.

Let me return here to something already suggested elsewhere [19]: “informatics” refers to “information”, but to what we colloquially call as “information” in biology, medicine, chemistry, physics, sociology, etc, is mostly as yet unknown and/or is too difficult to be currently defined with precision. In fact, defining information as “data + meaning” – an old idea [11], revisited with renewed enthusiasm [12, 20] – is just the result of another semantic interpretation, broadly adopted but still lacking any formal demonstration, and very subject to ongoing modification. In Elkin et al's manuscript – like in the AMIA and IMIA classifications, of course –, can we assume that we can precisely define BMI based only on the terms transmitted and used by authors/scientists/professionals/editors, even after being comprehensively grouped and assigned into taxonomies, classifications, terminologies, vocabularies, ontologies, etc? I think the answer is clearly: not yet, at least for many of us. Not everything is about semantics.

Despite, or because of the above comments – and digressions –, the manuscript of Elkin et al can be thought of as highlighting issues of long-term interest. While these types of papers can be soon outdated, their main value arises if their results are maintained and updated in frequent revisions, which can help a paper – and its associated taxonomy/ontology – to become an established reference in a field, particularly for academic and research purposes. As could happen in this case, of course.

8. Comment by F. J. Martin-Sanchez

In their article “Biomedical Informatics: We Are What We Publish” [1], Elkin et al. undertake an effort to characterize the field of biomedical informatics. They do this by analysing the articles in MEDLINE, which they retrieve by searching the literature using the terms ‘Medical Informatics’ and ‘Bioinformatics’. In order to do this, they use existing consensus classifications schemes endorsed by AMIA and IMIA. In addition, they attempt to enrich these re-

sults with findings from the literature obtained through several methods, such as natural language processing, ontology merging and parsing.

While the work described is valuable and must have been very time-consuming, I am doubtful about the usefulness of some of their results. The authors adequately point out some of the limitations of this study, such as not having used MeSH terms or the need to include semantic analysis of the literature instead of just concept mapping. However, I would argue that the article has conceptual and methodological weaknesses that require further discussion. These could be grouped into two categories: problems with the literature search and information processing; and questionable aspects associated with defining biomedical informatics using the scientific literature.

Literature Search and Information Processing

It is questionable that just searching for the terms ‘Bioinformatics’ and ‘Medical Informatics’ will ensure an adequate coverage of the scientific literature. Work in Bioinformatics has been also published under the labels of computational biology, systems biology, genome informatics and many other terms. Some of these articles would have not been identified in the search performed by the authors. In addition, terms like health informatics, nursing informatics or consumer informatics could have not been identified by the term “medical informatics”. Moreover, this exercise would have benefit from identifying articles published in the computing/IT literature databases (e.g. CiteSeerX) where relevant research work in this domain has also been published.

For a study aiming to determine coverage of medical informatics and bioinformatics attributable to the literature, the parsing procedure appears as somehow incomplete. For instance, within Bioinformatics, looking for gene names is not enough. Many proteins have different names to the gene that code for them. There is also published bioinformatic work on gene expression, metabolomics, and other areas that do not necessarily have to include gene names. Similarly,

in medical informatics, there is very valuable work in medical informatics dealing with health care delivery, public health or consumer health, to name a few, that is not necessarily annotated with disease names or body parts. Lastly, the size (27,000) of the random sample of articles used for the SNOMED parsing exercise does not seem to be well justified (why not the full set of articles?).

Given that the authors have parsed the literature with a new Health Informatics Ontology, which is the product of the manual merging between the AMIA consensus panel of core competencies and the IMIA classification of the field, the final result also includes competencies. An example is 'email' (in AMIA) or 'formulating questions' (in IMIA). I would argue that these concepts are of limited usefulness in order to define the scope of the field of biomedical informatics.

There have been some successful attempts to explore the scientific literature for emerging research topics in medical informatics and bioinformatics [21, 22]. Instead of searching for Medline articles with keywords, they started by creating a corpus of articles published in the main journals of these fields, followed by analysis of pairs of words (bigrams). In [23], the authors complemented the search with information from NIH-funded project grants and used MeSH Terms. In all these cases, the research teams were able to identify areas of rapid growth in bio- and medical informatics.

Defining Biomedical Informatics Using the Scientific Literature

Given the authors' claim that top-down approaches to modeling the biomedical informatics field are incomplete, as well as the problems described above, I would argue that this data-driven exercise based on the study of the literature is very limited, and therefore adds little value to the existing top-down classification schemes.

In this context the quotation by Cohen [24] is relevant: "*Defining a discipline through a keyword classification scheme of research is like driving while looking through the rear-view mirror.*"

Dr. Joe Lex expressed a similar idea at the International Emergency Medicine

Education Efforts & E-Learning Conference in 2012:

"If you want to know how we practiced medicine 5 years ago, read a textbook.

If you want to know how we practiced medicine 2 years ago, read a journal.

If you want to know how we practice medicine now, go to a (good) conference.

If you want to know how we will practice medicine in the future, listen in the hallways and use FOAM."^a

FOAM stands for Free Open Access Meducation. The term FOAM was coined in June 2012 in Dublin, during the International Conference on Emergency Medicine (ICEM). Many emergency physicians and intensivists based in Australia and New Zealand are now part of this movement that has spontaneously emerged from the increasing availability of collaborative, interactive and freely available resources for medical education being distributed on the web and social media.

Although this last quotation refers to the practice of medicine, I believe it is perfectly applicable to the practice of Biomedical informatics. I have no doubt about the importance of properly defining our discipline, this is a crucial step if we are to increase its recognition and improve the way educational programs are designed. However, it seems difficult to do this just looking into published research. Perhaps the analysis of social media channels and the scan of our conferences' content could provide us with a more accurate vision of the elements that make up the body of knowledge that supports our profession. For example, in the ontology resulting from the work of the authors (Health Informatics Ontology), there seems to be insufficient coverage of current issues relevant to biomedical informatics, such as clinical genomics, pharmacogenomics, data processing of individual exposure to environmental risk factors (exposome) or the role of informatics to support participatory medicine (social networks, self-monitoring, patient reported outcomes, mobile health).

Finally, when defining an academic discipline, it should be noted that there are

^a <http://lifeinthefastlane.com/foam/> (Last accessed 6 October 2013)

different ways to do this, for example, by its methods, concepts and theories, or by the object or subject of study. The current classification of Biomedical Informatics in application subdomains (translational bioinformatics, consumer health informatics, clinical research informatics, clinical informatics and public health informatics)^b seems more leant towards the second criterion. This classification has been very useful, as a way of integrating medical bioinformatics work involved in the development of genomics and personalized medicine in the years 2000. However, as argued by Edward Willatt on his Blog^c, defining a discipline by the first criterion has some advantages. "*Disciplines defined by their universal methods show themselves to be capable of realising genuine change as well as having the scope of the concrete and not simply of a special subject-matter.*"

In summary, I read with great interest the article by Elkin et al. The authors are to be complimented for providing an interesting approach to characterizing the biomedical informatics discipline using the scientific literature. This work, though, has to be considered just an initial step and much more work needs to follow. The methods used for the analysis of the literature must ensure a broader and deeper coverage of existing articles and learn from the previous experiences reported elsewhere. I also suggest that this analysis should include the scan of social media channels and biomedical informatics conferences content. Finally the article opens the debate on whether our discipline should be re-defined and how. Retrospective versus prospective, by subject-object versus by universal methods? These are important issues that will require further research.

9. Comment by L. A. Moura

Trying to define an evolving field is not an easy task. In the case of Health Informatics

^b <http://www.amia.org/about-amia/science-informatics> (Last accessed 6 October 2013)

^c <http://edwardwillatt.yolasite.com/blog/defining-a-discipline> (Last accessed 6 October 2013)

this is made even more difficult by its multidisciplinary nature and by the fact that the pace of change in Health Informatics is really fast, when compared to other areas, and is still on the increase.

"We are what we publish" is an interesting and provocative title (and a statement). Certainly, what we publish is part of what we are. Therefore, I see a great deal of value in the paper authored by Elkin, Brown and Wright [1]. Their methods are fine and their foundations solid. Their conclusions are logical and make sense.

Having said that, my eHealth practitioner self (as opposed to my academic self) will take the role of defending a different view.

My starting point is that – in a fast evolving field like ours – what we practice today is based on our past experience and on the current environment, but the most innovative deeds we do now will only be published in the future. It is also difficult, if not impossible, to keep up with tags or subject headings for classification as appropriate.

We can start by analyzing the past, focusing on terms that describe our field. More than forty years ago, IMIA was created under the "Medical Informatics" concept, which clearly showed to be insufficient to describe the field of work it intended to. Along the years, a plethora of terms have been used interchangeably to express the field of application of ICT techniques, methods and tools to Health and Health Care. Such terms include Health Informatics, Clinical Informatics, Health and BioMedical Informatics and, more recently, eHealth.

In my view, Health Informatics is the broadest, deepest, most stable and therefore most suitable single term to describe our field. The use of alternative terms is greatly dictated by the surrounding environment. Without looking for clear evidence, in my perception the term "Health and BioMedical Informatics" seems to have gained strength from a series of requests of proposals by the US government, which led research groups to adopt the term. Likewise, the now very prominent "eHealth" has surfed the waves of the Internet, following eBusiness and eGovern, her older siblings, to become so popular. eHealth is a

wonderful term, as it grasps the notion of agility associated with the prefix "e" and thus expresses the notion of greatly-connected and very efficient Health Systems. eHealth epitomizes all the goods that come from Health Informatics in a single concept that is simple, meaningful and powerful, at least for today. It should be expected, though, that – as happened with her older siblings – eHealth loses some strength as it becomes more widespread. When it turns out more intertwined with Health, eHealth is likely to become more transparent and thus invisible. The lifecycle for the term Health Informatics seems to be longer and the field it represents will not lose its meaning in the foreseeable future.

Coming back to the focus of the paper, IMIA needs to recognize that many of its members and partners describe our practice today as "eHealth". So much so that in 2011 IMIA and WHO organized and published an especial WHO Bulletin issue on evidences for eHealth [32]. In 2012, ITU and WHO published their National eHealth Strategy Toolkit [28], an exhaustive set of tools for helping countries achieve what came to be expressed as a resolution. In fact, with support from IMIA, the World Health General Assembly approved of its resolution 66.24 [33], in May 2013, urging member countries to develop and implement National eHealth Strategies. Of course, several countries – not confined to the First World anymore – have published their eHealth Strategies [25, 26, 29, 30].

More important than the current terms used to describe the Health Informatics field are the activities we perform today and that have seldom been done before. The volume of publications on strategies and organizational resources such as architectures associated to Health Informatics has increased in recent years. ISO TC 215 published, in 2012, its TR 14.639 on eHealth Architecture, aiming at Low and Middle Income Countries [27].

As Health Informatics has become more mature and Health Information Systems more mainstream, non-technical issues such as organizational matters – e.g. strategies, architectures, leadership, governance, policies and regulations – have started gaining due attention. However, although

sizzling and very relevant for the present and the future of Health Informatics, such activities have not been properly represented in the current study, as most of the keywords – eHealth, for one – are not included in the scope of the paper. In fact, eHealth doesn't even appear in the paper, as it is not part of the vocabularies and ontologies used. In a quick query to PubMed, I've just found 2,217 papers with "eHealth" or "e-Health" as an explicit keyword. Apart from that, other important eHealth initiatives may be in the making, and thus have not yet been published.

In my view, my comments point out to the need of updating IMIA Classification of the field of Biomedical Informatics and related documents.

My comments are not meant to and actually do not take away the merits of the paper, but I hope they raise the attention to some of its limitations, which are essentially due to trying to define this complex and ever-evolving field of ours. Health Informatics is still moving swiftly. Any attempt at providing a snapshot of it, using robust scientific methods, is likely to be hindered by the speed of its changes. Yet, we need to keep working on this definition as they are of foremost importance for establishing Health Informatics as a science and practice field, characterized by a body of knowledge of its own.

An important additional comment must be made regarding the term "mHealth", also very popular. For those in our field, mHealth tends to be regarded as a subset of Health Informatics. mHealth is very appealing as mobile technology is ubiquitous and represents empowerment to the user. However, mHealth can only unleash its potential as part of formal Health Informatics or eHealth environments which provide the platform that allows mHealth to be scalable and interoperate.

10. Comment by F. G. B. de Quirós

The article by Elkin and collaborators [1] has an interesting strategy in approaching a very significant topic of the discipline, which is its own definition and spectrum.

The research question aims to evaluate if AMIA's and IMIA's core competencies definitions describe in breadth and depth the characteristics of the discipline, by comparing them with the published literature indexed with MeSH terms "Medical Informatics" OR "Bioinformatics" for the year 2010. They analyzed 168,298 unique articles.

The authors defined an ontology based on the AMIA Consensus Panel Core Competencies for the Field of Biomedical Informatics [9] and the IMIA's Knowledge Base [12] creating their own AMIA-IMIA Health Informatics Ontology (HIO). It contained 433 health concepts and 462 terms which were built into a terminology server using an intelligent natural language processor (iNLP).

Titles and abstracts were parsed through the iNLP HIO, finding 62,244 of the 168,298 articles (37%), containing at least one associated concept. When combined with HGNC Ontology the coverage increased to 45.6%. Also a random sample of 27,000 articles with abstract was parsed with SNOMED CT to evaluate the overlap with clinical medical subspecialties and HGNC ontology for bioinformatics articles.

The authors describe some limitations. We would like to add some other comments and potential limitations.

The lack of lexical richness of the ontology (concepts and its synonyms) might bias the results. The authors define a total of 433 concepts with 462 terms, which seem to have few lexical variants and synonyms. This could cause a lesser retrieval when using iNLP, since the words used in the abstracts might have lexical variants that were not represented in the ontology.

When matching with SNOMED CT, so as to evaluate the relationship with medical specialties, they found 37,000 concepts that were grouped by "disorder by body site", and represent the findings and what was not retrieved in Tables 3 and 4. Children's of "Disorder by body site" in SNOMED CT includes a quite heterogeneous collection of concepts. The nature of a Description Logics (DL) classified ontology brings some unpredictability to this kind of top-down groupers, so these concepts are probably not a good representation of all body

sites or medical specialties [34]. Also this grouping might have left out procedures, drugs and devices. The authors could have grouped using the "body structure" hierarchy, that has a better anatomic organization, and being primitive concepts they are not re-organized based on automatic algorithms. This approach would also allow to group finding and procedures using the finding site or procedure site, attributes of the concept model.

Beyond these comments or potential limitations, the results are of great interest to evaluate the characteristics of the discipline. They propose that the literature might enrich the description of the breadth and depth of the content of our specialty.

The hypothesis about a gap between expert definitions and published experiences seems challenging. We could add to this hypothesis that there also might be a gap between what is published and what the "professionals in the field" actually do, considering that a number of daily activities carried out in the discipline don't get published due to lack of planning for proper evaluations, local limitations and barriers for research or because they take place in non-academic settings where publishing is not part of the culture [35, 36].

This situation creates a challenging question, partially suggested by the authors. **Which is the best way of defining what an "evolving" discipline is?**

In the relationship between the definition of the discipline and the daily activities, we can identify three levels:

1. Activities carried out by professionals on the field (**Facts**).
2. Discipline representation by experiences published in the literature (**Articles**).
3. Definition of the discipline by expert consensus (**Definitions**).

Although these relationships might be hierarchical and complete, i.e. the experts consider all the breadth and depth of what is published, and what is published includes all the activities carried out in the field, there are different reasons why we could lose breadth and depth between levels.

The "knowledge pyramid" of data, information and knowledge [37] represents

an analogy for this concept. Ackoff [38] published a definition of data, information and knowledge (DIK) stating that:

- "Data are symbols that represent properties of objects, events and their environment. They are the products of observation. But are of no use until they are in a useable (i.e. relevant) form. The difference between data and information is functional, not structural".
- "Information is contained in descriptions, answers to questions that begin with such words as who, what, when and how many. Information systems generate, store, retrieve and process data. Information is inferred from data." So information is data organized to answer questions.
- "Knowledge is know-how, and is what makes possible the transformation of information into decision making and instructions. Knowledge can be obtained either by transmission from another who has it, by instruction, or by extracting it from experience." Knowledge might be obtained analyzing information in context.

Ackoff also includes in this pyramid the wisdom level, as the dimension that represents the answer to the "why" question.

So the analogy would be DIK to FAD (facts – articles – definitions). Facts would represent data that is not usable for defining the discipline, but are the basis for answering the research questions that will transform these facts (data) into articles (information). The information will answer the "who, what, when and how many" questions, resulting in articles published on the literature. But the "know-how" to represent a discipline might need the expert to put it into context, providing knowledge. In this sense, Davenport and Pruzak [39] described knowledge as a "fluid mix of framed experience, values, contextual information, expert insight and grounded intuition that provides an environment and framework for evaluating and incorporating new experiences and information".

In summary: in our analogy the discipline could be represented by a series of events that take place in the field, which become publications when they answer a

research question. Then experts put these results into context creating definitions and knowledge. Also, the definitions modifies some publications and the publications some events in the ground. In other words, it is not a categorical dilemma between what is published and experts define. It seems to be a continuum and dynamic multidirectional interaction between facts, publications and consensus definitions.

Given a “snapshot” of this dynamic reality as author did, it seems reasonable to expect some “physiological” differences between components of the FAD continuum. Nevertheless although there is not a “relevance” analysis of the retrieve articles not recognized by iNLP HIO, the difference found is significant.

This difference might be bias or over-estimated because of chance or study limitations. It also might represent a real difference. In this latter case, another question might be: do we publish all what we do and do experts put all the depth and breadth of the publications into the context to define the discipline? This transforms the “conceptual” original research question of Elkin and colleagues about how might publications enrich the depth and breadth of expert definition of the discipline into a more “operational” one: is everyone (professionals, researchers, funders, universities, publishers, editors, reviewers, and experts) in the FAD continuum aligned in their objectives and playing his role at a high quality level?

11. Comment by M. J. Schuemie and P. Moorman

Although the paper by Elkin et al. [1] was interesting to read, it left us puzzled at the end; What have the authors actually studied, and what does it entail?

We think that the authors have shown that only a limited number of terms in the Health Informatics Ontology (HIO) are found in the biomedical informatics literature. What does this mean? Are the authors suggesting that the terms that were not found are possibly not relevant for our field, and should not be in the HIO if we

are what we publish? Or are they pointing out that we should start doing proper research in the direction of topics that are currently not covered? Even more: were there any topics found that were not part of HIO? Unfortunately, the authors do not provide us with answers to these questions.

Moreover, there are a few methodological issues.

The first deals with their search strategy. The authors describe their search as “Medical Informatics’ OR ‘Bioinformatics’”, which MEDLINE translates into “‘medical informatics’[All Fields] OR ‘bioinformatics’[All Fields]”. The importance of this observation lies in the fact that

- a) all research where one of the authors is part of a Medical Informatics Department (maybe combined with a Statistics or epidemiology part of that department) is considered to be on a medical informatics topic. A search with “medical informatics”[ad] makes it clear that that is questionable, and
- b) the strength of searching (also) with Mesh-terms and their underlying concepts has not been used. Thus, it seems that the retrieved papers dealt mostly with the broad Medical Informatics concept, but possibly not with more general concepts such as electronic medical records and electronic decision support. The fact that a search with (“medical records systems, computerized” OR “Decision Support Systems, Clinical”) NOT (“medical informatics” OR “Bioinformatics”) alone, already retrieves almost 20,000 references queries whether the authors search was indeed representative of the field.

A second methodological issue lies in the fact that the authors do not provide us with much insight into HIO, although Figures 1 and 2 give some hints. We wonder whether the software indeed searched for verbatim copies of the term “Data modeling to support analysis: warehouse/retrieval/EBP” in the literature? If so, then why do the authors not mention the fact that many articles could possibly discuss the subject (e.g. Ogunyemi et al. [40]), but using other phrases and words. Even though the authors claim that their software achieves very high precision and recall, we under-

stand that this performance was measured on a different task, namely the mapping of items in clinical problem lists to SNOMED CT [41]. A task that is difficult to compare to the task described in the paper.

A further point of concern is the described evaluation of overlap between our field and clinical medical subspecialties. First of all, the wording ‘overlap’ is awkwardly chosen. We believe that, in medical informatics research we apply our theories, methods and algorithms to medical domains. The question in that part of the studies’ evaluation therefore should rather be ‘what part of our research was done in an identifiable clinical domain?’, which would indeed be an interesting question. However, Figure 7 shows us that the iNLP server marked concepts such as “more accurate” and “changed”, which are indeed in SNOMED CT (e.g. “changed” is a quantitative concept), but are not necessarily medically relevant. These examples clearly show that not all SNOMED CT concepts indicate medical specialties.

Finally, the authors also study gene-coverage in the biomedical publications. Detecting which genes are mentioned in scientific literature is extremely difficult. Gene names are often abbreviations or acronyms which introduces many homonym and synonym problems in gene nomenclature [42]. For example, the HGNC contains the synonym “PSA” for the KLK3 gene, a term that is notoriously ambiguous [43] and can stand for many things, including “Poultry Science Association”. We therefore think that ‘gene mention detection’ is a field in its own right.

In conclusion, even though we agree with Elkin et al. that a bottom-up approach to defining the field of (bio)medical informatics may have value, we are not sure that the methodology applied here really achieves this. And although Elkin qualifies our previous research [13, 44] as top-down, we do not agree: First, instead of predefining the set of articles to include in our analysis, we started with a seed set of medical informatics journals that are clearly part of the field, and automatically added literature that was similar to this seed. Second, we used clustering of the selected papers and automatic identification of the words and combinations of words unique

to the clusters to identify key concepts. In this way, we let the underlying structure of the literature speak for itself. It would be interesting to see how these concepts could be related to the HIO.

12. Comment by B. Smith It Usually Begins with the Gene Ontology

Biomedical ontologies have now become a standard part of the biomedical informatician's toolkit. Initially, with the Gene Ontology (<http://geneontology.org>), ontologies were introduced to enhance the comparability of gene array data deriving from research on different model organisms. Very rapidly they began to be used as tools to enhance the discoverability of data more generally, to allow new sorts of statistical analyses of data under the heading of ‘gene enrichment studies’ and to allow the merger of large bodies of data deriving from different sources through the use of a common set of ontology annotations.

This latter application is of increasing importance especially in translational medicine and in interdisciplinary areas such as research on aging, where ontologies are playing what we might think of as an educational role. In aging research, for example, researchers working on nutrient sensing might be called upon to collaborate with those working on mitochondrial dysfunction, or on stem cell exhaustion, and all of these might in turn need to collaborate with experimentalists working on apoptosis in yeast. To a surprising degree the Gene Ontology is serving as a common resource upon which all of these communities are able to draw in combining their data. I believe that part of what is going on here is that, when human beings need to formulate and test hypotheses and to display and analyze experimental results involving contributions from unfamiliar disciplines, the GO is used, in effect, as a simple educational aid.

But the FMA Is also Involved

In fact researchers in biomedical ontology already from the very start have been suggesting that ontologies might serve an edu-

cational role of an even more ambitious sort. Cornelius Rosse and his collaborators in Seattle proposed as early as 1998 that ontologies could serve as a platform to re-engineer education in the core basic sciences. The discipline of anatomy, as they pointed out

is the first subject – and one of the most challenging and time-consuming subjects – introduced in the training of all health care professionals. There is a need for logic-based, machine-parsed representations of anatomical knowledge for the creation of intelligent educational programs in anatomy.

What they at that stage referred to as the ‘Digital Anatomist ontology’ and has since been transformed into the Foundational Model of Anatomy Ontology [45, 46] would, they held, establish ‘a basic requirement for such applications’ and would ‘serve as a platform for a digitally re-imagined approach to the teaching of anatomy as core basic science in medical education programs’ [47].

This ‘digitally re-imagined approach’ would then be applied not merely through the FMA anatomy reference ontology but also through reference ontologies in other areas of basic science, including genetics, cell biology, physiology, and so forth. A single set of reference ontologies would in this way – given the full realization of Rosse's vision – become engrained in the course of medical training on the very brains of medical students. These ontologies would then automatically work in tandem with the ontologies being used to capture the clinical data which these medical students are using in their daily activities, since the latter would be built up on the basis of the former.

We Are What We Publish

Elkin, Brown and Wright, in their “Biomedical Informatics: We Are What We Publish” [1], formulate what we can think of as an ambitious complement to Rosse's vision. They argue in effect that we can not merely use ontologies as a vital tool of medical education, but that we can go further and use the ontological approach to determine the very content of one (and not the least important) branch of the biomedical curriculum. They make this proposal in

the context of an analysis of the AMIA and IMIA initiatives to formalize the definition of ‘biomedical informatics,’ extracting to this end the terms used in the AMIA consensus document and combining these with the terms employed in the IMIA definitions. They then built manually on this basis a draft Health Informatics Ontology, which they used to parse a very large corpus of medical literature identified using NLP software, with “Medical Informatics OR Bioinformatics” as search criterion.

The results are of interest from a number of different points of view. But they show that the merged AMIA-IMIA-based ontology is able to identify the coverage domain of biomedical informatics only partially, in that of the 168,298 articles identified, only some 37% contained at least one term from the HIO in its title or abstract. Work is accordingly on-going on a new version of the HIO, both expanded and more formal, in order to establish the degree to which there is material published in the field of biomedical informatics that is not covered by the AMIA/IMIA specifications.

Such an expanded HIO could then be used for more ambitious investigations – for example to provide a series of snapshots of the discipline to demonstrate how it has changed, and is still changing, over time. The enhanced ontology would contribute, as the authors point out, to a greater self-understanding of the discipline of biomedical informatics by its practitioners – and it could thereby also help to realize the vision for ontology as a tool for biomedical informatics education along the lines proposed by Rosse.

At the same time, however, we can see some of the problems facing such a vision. As the authors acknowledge, the HIO itself is still in early draft stage, and it lacks formal definitions of its constituent terms. It was moreover developed on the basis of inputs created through both an AMIA and an IMIA consensus process that was not aimed at yielding an ontological representation of a principled sort. The result requires work to adapt the HIO to best practice principles for ontology development, including those identified through the OBO Foundry initiative (<http://obofoundry.org>).

To be of value to the process of biomedical education, integration of HIO with the reference ontologies corresponding to the basic biomedical sciences would also be important. Building ontologies using what the authors call 'concepts' in the biomedical literature and relying on the HIO to provide semantic context along the lines the authors propose will yield satisfactory results only if the HIO itself is in good shape from an ontologico-semantic point of view, and for this considerable further effort is needed. The results should then satisfy not merely consensus review by the practitioners of the specialty of biomedical informatics, but also survive stringent examination by specialists in the field of ontology.

Creating an HIO in this manner will be no easy task. In contrast to human anatomy, which is an evolutionarily highly stable domain marked by a considerable degree of disciplinary self-understanding, biomedical informatics is an inherently complex and interdisciplinary and above all dynamically evolving field. As the GO has shown, an ontology can demonstrate considerable practical value even in a rapidly changing field of scientific endeavor. Having taken it upon themselves to create the Health Informatics Ontology, the authors now have the responsibility to work with the ontology community to demonstrate that they can together create an artifact marked by the sort of ontological rigor that would make it truly useful in defining and shaping the field of biomedical informatics.

13. Comment by J. Talmon

The title of the Elkin et al. paper [1] raises high expectations. One would expect to find a description of our field based on what we

have published. At the end of the manuscript we are still (partially) in the dark.

As a frame of reference, the authors used the Health Informatics Ontology (HIO). Their main finding is that only 37% of the retrieved articles had a concept in the title or abstract that also occurred in the HIO. On the other hand only 251 of the 433 concepts of the HIO could be identified in the literature. One would have expected that a deeper analysis was made of the articles that did not have a concept of the HIO as well as of the concepts in the HIO that were not found in the literature. Such an analysis would have revealed were the discrepancy is and whether or not we should revise the top-down developed AMIA and IMIA terminologies. At least some supplementary material should have been provided to allow the reader to better understand these discrepancies.

The analysis of the random sample of 27,000 articles with abstracts by parsing them with SNOMED-CT seems too much focused on disorders rather than clinical subspecialties. Unfortunately there is no data on the number of articles that did not have a SNOMED-CT code for a disorder by body site.

Apart from the disappointment related to the limited analyses and the rather general discussion, there are a few methodological issues I would like to raise.

A main concern is the method used to define what is being published in our field. In the Schuemie et al. paper [13] we started from what is being published in the journals that are defined by Thomson Reuters to cover the field of Medical Informatics; indeed we did not consider bioinformatics. From there we identified which 1-, 2-, and 3-grams were more common in the corpus of MI publications as compared to the rest

of the PUBMED corpus. We also investigated whether there are other series or journals indexed in PUBMED that also published in our domain. As a matter of fact we tried to identify how our field differentiates itself from other disciplines, not on what we may have in common.

Elkin et al. on the other hand relied on two search terms to retrieve articles from PUBMED. One should be aware that searching *Medical Informatics* is different from searching for "Medical Informatics". The former resulted in 153,403 hits, the latter, however, in only 9172 hits^d. This makes it clear that what you search for will influence the results, and thereby what our domain entails.

It is strange that only *Medical Informatics* and *Bioinformatics* have been used as search terms. In their paper, Elkin et al. also use the term biomedical informatics. This term gives 1815 hits in PUBMED, 260 of which were not covered by searches for *Medical Informatics* or *Bioinformatics*.

It seems that large areas of application of ICT in health care have not been covered. For example, Telemedicine – 8408 hits of which 5332 were not covered by the search for *Medical Informatics* or *Bioinformatics* – is hardly dealt with.

A further concern is that not all of what we publish in our Medical Informatics Journals has been retrieved by the queries of Elkin et al. IJMI, MIM, JAMIA and JBI have published 4103 articles prior to February 2006. Of those only 2871 appeared in queries for *medical informatics* or *bioinformatics*. Nearly one third of what is being published in our journals is not accounted for. Are all those papers outside the domain of (Bio)medical Informatics?

A final note is on the time span of the search used by Elkin et al. They performed their search in February 2006. We are now more than seven years later. In that time, the body of literature on (bio)medical Informatics has more than doubled (► Table 1). In particular given the large increases in the number of publications in bio(medical)informatics, this raises the question how valid the findings still are.

^d All PUBMED searches have been done for publications prior to 01/02/2006.

Table 1 Number of publications at different time instances of search and their percentage increase

Query	Before February 2006	Before November 2013	Percentage increase
Medical Informatics	153,403	307,700	100%
Bioinformatics	32,238	125,655	290%
Telemedicine	8,408	16,254	93%
"our journals"	4,103	7,457	81%
Biomedical Informatics	1,815	6,492	257%

Given these concerns about the search strategy, one can question whether this study has really studied what we have published. The current paper presents results of the application of a set of useful NLP tools on some selection of papers from PUBMED, which is not necessarily representative for what we publish in our field. Hence the external validity of the results is limited.

References

- Elkin PL, Wright G, Brown SH. Biomedical Informatics: We Are What We Publish. *Methods Inf Med* 2013; 52 (6): 538–546.
- Miller RA, Masarie FE Jr. The demise of the “Greek Oracle” model for medical diagnostic systems. *Methods Inf Med* 1990; 29 (1): 1–2.
- Carroll L. Through the Looking-Glass, and what Alice Found There. Chapter VI. London, United Kingdom: Macmillan and Co.; 1872.
- Ammenwerth E, Schreier G, Hayn D. Health Informatics Meets eHealth. *Methods Inf Med* 2010; 49 (3): 269–270.
- Rockefeller Foundation. Solutions from Silos to Systems: An Overview of eHealth’s Transformative Power. New York (USA): Rockefeller Foundation; 2010.
- Global Observatory for eHealth. ATLAS: eHealth Country Profiles (2010), based on the findings of the second global survey on eHealth. Geneva (Switzerland): World Health Organization; 2011.
- World Health Organization and International Telecommunication Union. National eHealth Strategy Toolkit. Geneva (Switzerland): World Health Organization; 2012.
- Mantas J, Ammenwerth E, Demiris G, Hasman A, Haux R, Hersh W, Hovenga E, Lun KC, Marin H, Martin-Sanchez F, Wright G; IMIA Recommendations on Education Task Force. Recommendations of the International Medical Informatics Association (IMIA) on Education in Biomedical and Health Informatics. First Revision. *Methods Inf Med* 2010; 49 (2): 105–120.
- Kulikowski CA, Shortliffe EH, Currie LM, Elkin PL, Hunter LE, Johnson TR, Kalet IJ, Lenert LA, Musen MA, Ozbolt JG, Smith JW, Tarczy-Hornoch PZ, Williamson JJ. AMIA Board white paper: definition of biomedical informatics and specification of core competencies for graduate education in the discipline. *J Am Med Inform Assoc* 2012; 19 (6): 931–938.
- Musen MA. Medical informatics: searching for underlying components. *Methods Inf Med* 2002; 41 (1): 12–19.
- Greenes RA, Shortliffe EH. Medical informatics. An emerging academic discipline and institutional priority. *JAMA* 1990; 263 (8): 1114–1120.
- Wright G. The Development of the IMIA Knowledge Base. *SA Journal of Information Management*. 2011; 13 (1), Art. # 458, 5 pages.
- Schuemie MJ, Talmon JL, Moorman PW, Kors JA. Mapping the domain of medical informatics. *Methods Inf Med* 2009; 48 (1): 76–83.
- Lok C. Literature mining: Speed reading. *Nature* 2010; 463 (7280): 416–418.
- Maojo V, Kulikowski C. Medical informatics and bioinformatics: integration or evolution through scientific crises? *Methods Inf Med* 2006; 45 (5): 474–482.
- Kulikowski CA. The micro-macro spectrum of medical informatics challenges: from molecular medicine to transforming health care in a globalizing society. *Methods Inf Med* 2002; 41 (1): 20–24.
- McCray AT, Gefeller O, Aronsky D, Leong TY, Sarkar IN, Bergemann D, Lindberg DA, van Bommel JH, Haux R. The birth and evolution of a discipline devoted to information in biomedicine and health care. As reflected in its longest running journal. *Methods Inf Med* 2011; 50 (6): 491–507.
- Haux R. Health care in the information society: what should be the role of medical informatics? *Methods Inf Med* 2002; 41 (1): 31–35.
- Maojo V, Kulikowski CA. Note on Friedman’s ‘what informatics is and isn’t’. *J Am Med Inform Assoc* 2013 Apr 16.
- Bernstam EV, Smith JW, Johnson TR. What is biomedical informatics? *J Biomed Inform* 2010; 43: 104e10.
- Rebholz-Schuhmann D, Martin-Sanchez F, Milanese L, Tollis I, Van der Lei J. SYMBIOmatics: Synergies in medical informatics and bioinformatics-exploring current scientific literature for emerging topics. *BMC Bioinformatics* 2007; 8 (Suppl): S18.
- Bansard JY, Rebholz-Schuhmann D, Cameron G, Clark D, van Mulligen E, Beltrame F, del Hoyo Barbolla E, Martin-Sanchez F, Milanese L, Tollis I, Van der Lei J, Coatrieux JL. Medical informatics and Bioinformatics: A bibliometric study. *IEEE Trans Inf Technol Biomed* 2007; 11 (3): 237–243.
- Perez-Iratxeta C, Andrade-Navarro MA, Wren JD. Evolving research trends in bioinformatics. *Brief Bioinform* 2007; 8 (2): 88–95.
- Cohen E. Reconceptualizing Information Systems as a Field of the Transdiscipline Informing Science: From Ugly Duckling to Swan. *Journal of Computing and Information Technology* 1999; 7 (3): 213–219.
- Council of Australian Governments (COAG). Heads of Agreement – National Health Reform. Available at http://www.coag.gov.au/coag_meeting_outcomes/2011-02-13/docs/communique_attachment_20110213.pdf
- European Commission. European countries on their journey towards national eHealth infrastructures. Final European progress report. 2011. Available at http://www.ehealth-strategies.eu/report/eHealth_Strategies_Final_Report_Web.pdf
- ISO TR 14.639-1:2012 Capacity-based eHealth Architecture Roadmap.
- National eHealth strategy toolkit. http://www.itu.int/pub/D-STR-E_HEALTH.05-2012. ISBN 978 92 4 154846 5 (WHO) (NLM classification: W 26.5). ISBN 978 92 61 14051 9 (ITU)
- South Africa – eHealth Strategy South Africa, 2012 – 2016. Available at http://www.doh.gov.za/docs/stratdocs/2012/eHealth_Strategy_South_Africa_2012-2016.pdf
- NHS Scotland – eHealth Strategy, 2011–2017. Available at <http://www.scotland.gov.uk/Resource/Doc/357616/0120849.pdf>
- Building foundations for eHealth: progress of Member States: report of the WHO Global Observatory for eHealth. Geneva, World Health Organization, 2006. Available at www.who.int/goe/publications.
- Bulletin of the World Health Organization. Volume 90, Number 5, May 2012, 321–400. Available at <http://www.who.int/bulletin/volumes/90/5/en/index.html>
- WHO – WHA Resolution. Resolution A66.24 on eHealth Standardization and Interoperability. Available at http://apps.who.int/gb/ebwha/pdf_files/WHA66/A66_R24-en.pdf
- Markwell D. SNOMED CT Technical Implementation Guide: International Health Terminology Standards Development Organisation; 2012. Available from: http://www.ihtsdo.org/fileadmin/user_upload/doc/en_us/tig.html?t=hier_body_struct.
- Canosa D, Ferrero F, Melamud A, Otero PD, Me-rech RS, Ceriani Cernadas JM. [Full-text publication of abstracts presented at the 33th Argentinean pediatric meeting and non publication related factors]. *Arch Argent Pediatr* 2011; 109 (1): 56–59. PubMed PMID: 21283945.
- Scherer RW, Langenberg P, von Elm E. Full publication of results initially presented in abstracts. *Cochrane Database Syst Rev* 2007; (2): MR000005. PubMed PMID: 17443628.
- Henry N. Knowledge management: a new concern for public administration. *Public Administration Review* 1974; 34 (3): 189–196.
- Ackoff R. From data to wisdom. *Journal of Applied Systems Analysis* 1989; 16: 3–9.
- Davenport T, Pruzak L. Working knowledge: How organizations manage what they know. Harvard Business Press; 2000.
- Ogunyemi OI, Meeker D, Kim HE, Ashish N, Farzaneh S and Boxwala A. Identifying appropriate reference data models for comparative effectiveness research (CER) studies based on data from clinical information systems. *Medical Care* 2013; 51 (8 Suppl 3): S45–52.
- Elkin PL, Brown SH, Husser CS, et al. Evaluation of the content coverage of SNOMED CT: ability of SNOMED clinical terms to represent clinical problem lists. *Mayo Clin Proc* 2006; 81 (6): 741–748.
- Morgan AA, Lu Z, Wang X, et al. Overview of Bio-Creative II gene normalization. *Genome Biol* 2008; 9 (Suppl 2): S3.
- Weeber M, Schijvenaars BJ, Van Mulligen EM, et al. Ambiguity of human gene symbols in Locus-Link and MEDLINE: creating an inventory and a disambiguation test collection. *AMIA Annu Symp Proc* 2003. pp 704–708.
- Moorman PW, Schuemie MJ and van der Lei J. An inventory of publications on electronic medical records revisited. *Methods Inf Med* 2009; 48 (5): 454–458.
- Rosse C, Mejino JLV. The Foundational Model of Anatomy Ontology. In: Burger A, Davidson D, Baldock R, editors, *Anatomy Ontologies for Bioinformatics: Principles and Practice*, 2007, volume 6, pp 59–117. London, Springer.
- Rosse C, Mejino JLV, Modayur BR, Jakobovits R, Hinshaw KP, Brinkley JE. Motivation and Organizational Principles for Anatomical Knowledge Representation. *The Digital Anatomist Symbolic Knowledge Base*. *J Am Med Inform Assoc* 1998; 5 (1): 17–40.
- Rosse C, Mejino JLV. A reference ontology for biomedical informatics: the Foundational Model of Anatomy. *J Biomed Inform*. 2003; 36 (6): 478–500.

Addresses of the Authors

Antoine Geissbuhler
 Department of Radiology and Medical Informatics
 Geneva University
 Rue Gabrielle-Perret-Gentil 4
 1211 Genève 14
 Switzerland
 E-mail: Antoine.Geissbuhler@hcuge.ch

William Ed Hammond
 Duke Center for Health Informatics
 2424 Erwin Road, Room 12053
 Durham, NC 27705
 USA
 E-mail: william.hammond@duke.edu

Arie Hasman
 Department of Medical Informatics
 AMC-University of Amsterdam
 Meibergdreef 15
 1105 AZ Amsterdam
 The Netherlands
 E-mail: a.hasman@amc.uva.nl

Rada Hussein
 The Biomedical Informatics Center of Excellence
 Information Technology Institute
 Ministry of Communications and Information Technology
 Smart Village, B 148
 28 Km Cairo-Alex Desert Road
 Giza
 PO Box 12577
 Egypt
 E-mail: rahussein@iti.gov.eg

Ross Koppel
 Sociology Department and the School of Medicine
 University of Pennsylvania
 553 McNeil Building
 Philadelphia, Pennsylvania 19104-6299
 USA
 E-mail: rkoppel@sas.upenn.edu

Casimir Kulikowski
 Department of Computer Science
 Rutgers, The State University of New Jersey
 Hill Center, Busch Campus
 Piscataway, NJ 08855
 USA
 E-mail: kulikows@cs.rutgers.edu

Victor Maojo
 Departamento de Inteligencia Artificial
 Facultad de Informática
 Universidad Politécnica de Madrid
 28660 Boadilla del Monte, Madrid
 Spain
 E-mail: vmaojo@infomed.dia.fi.upm.es

Fernando J. Martin-Sanchez
 Health and Biomedical Informatics Centre (HABIC)
 The University of Melbourne
 Level 1, 202 Berkeley Street
 3010 VIC.
 Australia
 E-mail: fjms@unimelb.edu.au

Peter Moorman
 Department of Medical Informatics
 Erasmus University Medical Center
 Rotterdam
 The Netherlands
 E-mail: p.moorman@erasmus.nl

Lincoln Assis de Moura
 Assis Moura eHealth
 Rua Miguel Tostes, 230
 90430-060 Porto Alegre, RS
 Brazil
 E-mail: lamoura@uol.com.br

Fernán Gonzalez Bernaldo de Quirós
 Department of Health Informatics
 Hospital Italiano de Buenos Aires
 Peron 4190
 (1199) Ciudad Autonoma de Buenos Aires
 Argentina
 E-mail:
 fernan.quirós@hospitalitaliano.org.ar

Martijn J. Schuemie
 Janssen Research & Development
 1125 Trenton Harbourton Road
 Titusville, NJ 08560
 USA