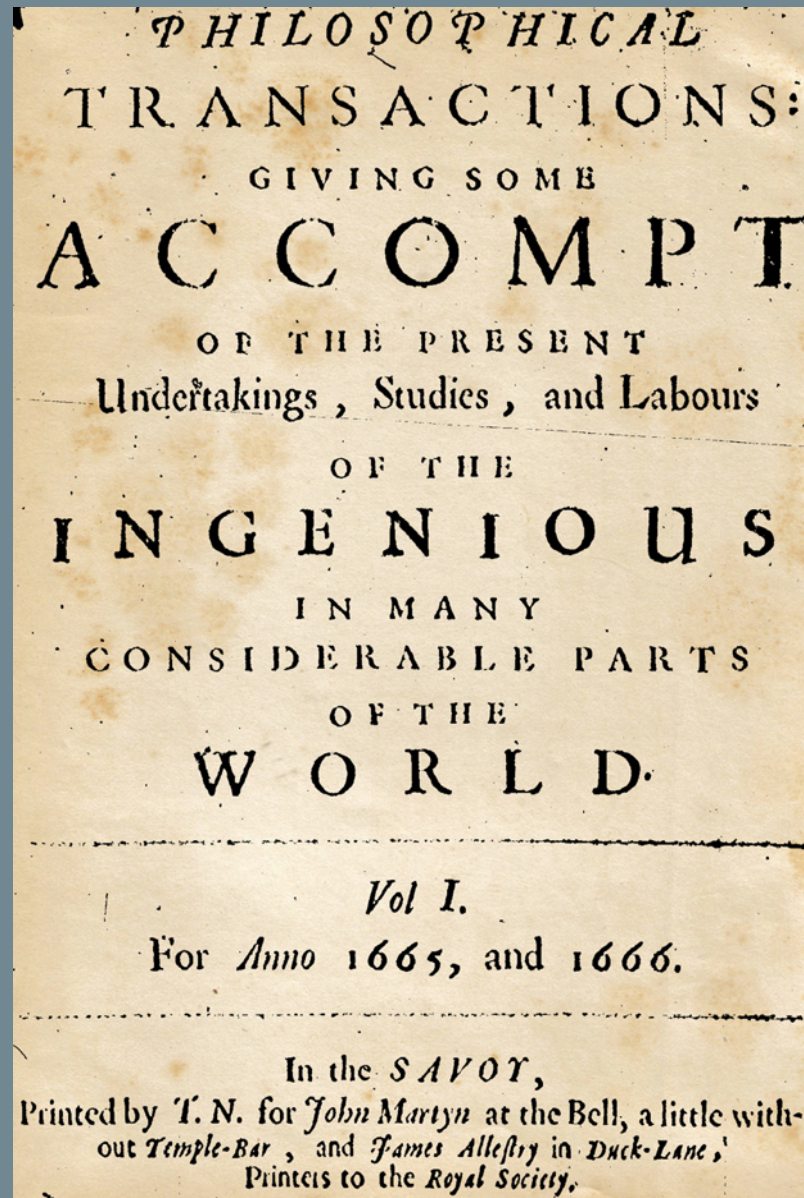


MTA SZTAKI DSD

Department of
Distributed
Systems

Kérdések ... avagy tudomány a 21. században

Kovács László



- posztakademikus tudomány
 - ipari - akadémiai kutatás
 - felfedező - alkalmazott kutatás
 - publikálás költségei
 - tudomány támogatása
 - tudományirányítás
 - tudományos közösségek szerepe
- papír túltermelés
 - tudományos szemét
 - reprodukálhatatlanság
 - plágium
 - csalások a tudományban
 - vállalkozás a tudományban

C

COMMUNALISM

the common ownership of scientific discoveries, according to which scientists give up intellectual property rights in exchange for recognition and esteem

U

UNIVERSALISM

claims to truth are evaluated in terms of universal or impersonal criteria

D

DISINTERESTEDNESS

scientists are rewarded for acting in ways that outwardly appear to be selfless

O

ORIGINALITY

novelty in research contributions

S

SCPTICISM

all ideas must be tested and are subject to rigorous, structured community examination

**PRINCIPLES OF “GOOD SCIENCE” - NORMS OF ROBERT K. MERTON
(THE NORMATIVE STRUCTURE OF SCIENCE, 1942)**

Mode-2 of Knowledge Production

started during the mid of 20th century, rapid transient during the last 20 years (“mode-2”- the new production of knowledge, 1994, Gibbons, Limoges, Nowotny, Schwartzman, Scott, Trow)

Context Driven

research carried out in a context of application, arising from the work of problem solving and not governed by the paradigms of traditional disciplines of knowledge, social and/or commercial interest, ...

Problem-Focused

Project based, research proposals, pragmatic finalisation is favoured above explanatory unification, technology as result, focus on reliability of knowledge in specific application, no fear of possible inconsistencies, ...

Interdisciplinary

transdisciplinary research, new type of creativity within the intellectual process, multidisciplinary research teams, large number of authors per papers, ...

Collaboration, Communication Driven

global collaboration, different time zones, internet-based communication, ...

ACADEMIC

SCIENCE



ROBERT K. MERTON

INDUSTRIAL



JOHN ZIMAN

- tudomány mint instrument (eszköz)
- elektronikus tudomány (eScience)
- hipotézis vezérelt tudomány - adatvezérelt tudomány
- alapadatok (raw data) és alapanyagok nyílt közösségi hozzáférésének biztosítása mint a publikálás előfeltétele (Nature)
- új minőségellenőrzési mechanizmusok (pl. crowdsourcing)
- open access
- szemantikus publikálás (cikkek szemantikus feltárása és publikálás a szemantikus web felé)

Editorial

Message from the Vice President of Publications on New Developments in Signal Processing Society Publications

THIS is a time of rapid changes in the publication world, with the opportunities that electronic publishing generates. Taking advantage of these changes, the IEEE Signal Processing Society staff and volunteers are looking for ways to better serve their authors and readers. In addition, the IEEE is also more broadly pursuing innovations in content delivery. I would therefore like to alert you to the advances that are already in place and developments to look for in the future.

Electronic publishing through IEEE Xplore offers the possibility of augmenting papers with multimedia examples—including audio, image, and video files—which provide concrete illustrations of the utility of a new algorithm. I encourage you to take advantage of these opportunities, which are especially relevant to many problems in signal processing and can increase the impact of your work. To make your work more easily reproducible, you are also encouraged to submit code that can be accessed through IEEE Xplore with your paper. IEEE Xplore is also now augmenting the paper presentation with links to related content and citation information, and more developments are in the works to make the presentation more useful to readers.

The Signal Processing Society has moved to an all-electronic version of the Content Gazette, with which you can electronically browse all our journals and click on links that take you straight to the papers that you are interested in. We welcome your suggestions of how to make the Content Gazette even more useful to you.

Electronic publication is also making it possible to reduce the time to move manuscripts from submission to publication, which again contributes to higher impact factors. In 2013, all of the Signal Processing Society journals will move to article-based electronic publishing, which allows a more frequent publishing schedule. Compiled print versions will be issued less frequently. We are also working with IEEE to reduce the editing time in general, referring papers to an outside editing service for the small percentage of cases in which more careful editing will be necessary.

Of course, timely reviews and revisions are still key elements to rapid publication! Part of moving papers through the publication process more quickly is timely notification to authors when a paper is not a good fit to the journal.

To assist authors and editors, we have updated our guidelines to establish explicit criteria by which papers can be immediately rejected without the full review process. In all such cases, the journal Editor-in-Chief gets input from the Editorial Board members, so there are multiple people considering every submission. By reducing the number of papers sent to full review, we hope to reduce the burden on reviewers and improve the quality of reviews overall.

Of course, one of the major developments right now is open access, which will be playing a larger role in the future of our publications. The IEEE already has in place a policy whereby authors can pay to have their paper available through open access within all our regular journals, but few people have taken advantage of this due to the high cost. Starting this fall, the cost will be lowered to \$1750 for a standard-length paper, which is in the range of other high-quality publication venues. You can take advantage of this option in any of our journals. The IEEE is providing other models for open access as well, which we are currently examining.

Lastly, we recognize the changing role of conference publications in our society. Conference publications are now easily available through IEEE Xplore, so they can no longer be thought of as “non-archival.” With this development, we need to rethink the relation between journal papers and conference papers, and we have therefore initiated a few changes.

First, we will no longer be accepting technical correspondence papers to our journals. Papers that are well suited to the four-page format should be submitted to the IEEE SIGNAL PROCESSING LETTERS or to one of our conferences.

Second, to encourage you to submit your best short-paper work to the IEEE SIGNAL PROCESSING LETTERS, we have instituted an agreement that authors of papers accepted to the IEEE SIGNAL PROCESSING LETTERS will have the option of presenting their papers in the next ICASSP or ICIP, if they so wish. This will reduce duplicate publication and allow quick turnaround for hot new ideas, while still allowing for the useful feedback associated with a conference presentation. Thanks to Anna Scaglione, Editor-in-Chief of the IEEE SIGNAL PROCESSING LETTERS, for this innovation and to the Conference Board for working with us on the implementation.

Finally, I want to remind authors and reviewers that we still accept articles that are extensions of work that has appeared in a conference publication, but the authors should make it clear what the extensions are in the submitted manuscript, which

with the opportunities that electronic publishing generates. Taking advantage of these changes, the IEEE Signal Processing Society's volunteers are looking for ways to better serve their authors and readers. In addition, the IEEE is also more broadly pursuing innovations in content delivery. I would therefore like to alert you to the advances that are already in place and developments to look for in the future.

Electronic publishing through IEEE Xplore offers the possibility of augmenting papers with multimedia examples—including audio, image, and video files—which provide concrete illustrations of the utility of a new algorithm.

I encourage you to take advantage of these opportunities, which are especially relevant to many problems in signal processing and can increase the impact of your work. To make your work more easily reproducible, you are also encouraged to submit code that can be accessed through IEEE Xplore with your paper. IEEE Xplore is also now augmenting the paper presentation with links to related content and citation information, and more developments are in the works to make the presentation more useful to readers.

The Signal Processing Society has moved to an all-electronic version of the Content Gazette, with which you can electronically browse all our journals and click on links that take you straight to the papers that you are interested in. We welcome your suggestions of how to make the Content Gazette even more useful to you.

Electronic publication is also making it possible to reduce the time to move manuscripts from submission to publication, which again contributes to higher impact factors. In 2013, all of the Signal Processing Society journals will move to article-based electronic publishing, which allows a more frequent publishing schedule. Compiled print versions will be issued less frequently. We are also working with IEEE to reduce the editing time in general, referring papers to an outside editing service for the small percentage of cases in which more careful editing will be necessary.

Of course, timely reviews and revisions are still key elements to rapid publication! Part of moving papers through the publication process more quickly is timely notification to authors when a paper is not a good fit to the journal.

to establish explicit criteria by which papers can be immediately rejected without the full review process. In all such cases, the journal Editor-in-Chief gets input from the Editorial Board members, so there are multiple people considering every submission. By reducing the number of papers sent to full review, we hope to reduce the burden on reviewers and improve the quality of reviews overall.

Of course, one of the major developments right now is open access, which will be playing a larger role in the future of our publications. The IEEE already has in place a policy whereby authors can pay to have their paper available through open access within all our regular journals, but few people have taken advantage of this due to the high cost. Starting this fall, the cost will be lowered to \$1750 for a standard-length paper, which is in the range of other high-quality publication venues. You can take advantage of this option in any of our journals. The IEEE is providing other models for open access as well, which we are currently examining.

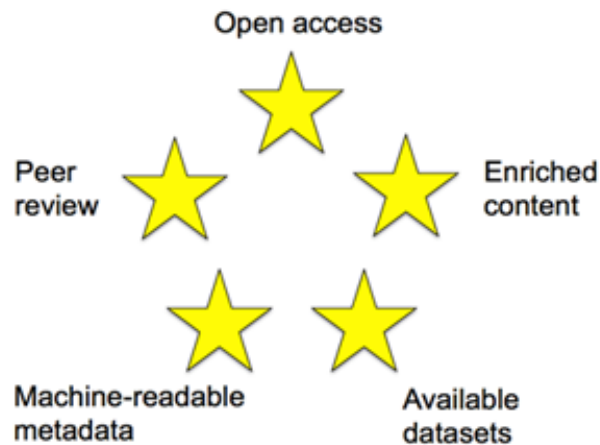
Lastly, we recognize the changing role of conference publications in our society. Conference publications are now easily available through IEEE Xplore, so they can no longer be thought of as “non-archival.” With this development, we need to rethink the relation between journal papers and conference papers, and we have therefore initiated a few changes.

First, we will no longer be accepting technical correspondence papers to our journals. Papers that are well suited to the four-page format should be submitted to the IEEE SIGNAL PROCESSING LETTERS or to one of our conferences.

Second, to encourage you to submit your best short-paper work to the IEEE SIGNAL PROCESSING LETTERS, we have instituted an agreement that authors of papers accepted to the IEEE SIGNAL PROCESSING LETTERS will have the option of presenting their papers in the next ICASSP or ICIP, if they so wish. This will reduce duplicate publication and allow quick turnaround for hot new ideas, while still allowing for the useful feedback associated with a conference presentation. Thanks to Anna Scaglione, Editor-in-Chief of the IEEE SIGNAL PROCESSING LETTERS, for this innovation and to the Conference Board for working with us on the implementation.

Finally, I want to remind authors and reviewers that we still

The Five Stars of Online Journal Articles



Peer review

Ensure your article is peer reviewed, to provide assurance of its scholarly value, quality and integrity.

Open access

Ensure others have cost-free open access both to read and to reuse your published article, to ensure its greatest possible readership and usefulness.

Enriched content

Use the full potential of Web technologies and Web standards to provide interactivity and semantic enrichment to the content of your online article.

Available datasets

Ensure that all the data supporting the results you report are published under an open license, with sufficient metadata to enable their re-interpretation and reuse.

Machine-readable metadata

Publish machine-readable metadata describing both your article and your cited references, so that these descriptions can be discovered and reused automatically.

David Shotton (University of Oxford):

The Five Stars of Online Journal Articles – a Framework for Article Evaluation

(D-Lib Magazine, Vol. 18., No. 2. January 2012)

Utopia dokumentum formátum és kezelő
szoftver

BioChemical Journal (www.biochemj.org)

Bj www.biochemj.org

Biochem. J. (2009) 424, 317–323. Printed in Great Britain. doi:10.1042/BJ20091474 317

REVIEW ARTICLE

Calling International Rescue: knowledge lost in literature and data landslide!

Teresa K. ATTWOOD*¹, Douglas B. KELL², Philip McDERMOTT*¹, James MARSH², Steve R. PETTIFER*³ and David THORNE⁴

¹School of Computer Science, The University of Manchester, Oxford Road, Manchester M13 9PL, U.K., ²Faculty of Life Sciences, The University of Manchester, Oxford Road, Manchester M13 9PL, U.K., ³School of Chemistry, The University of Manchester, Oxford Road, Manchester M13 9PL, U.K., and ⁴Manchester Interdisciplinary Biocentre, 201 Princess Street, Manchester M1 7DN, U.K.

We live in interesting times. Portents of impending catastrophe pervade the literature, calling us to action in the face of unmanageable volumes of scientific data. But it isn't so much data generation *per se*, but the systematic burial of the knowledge embedded in those data that poses the problem: there is so much information available that we simply no longer know what we know, and finding what we want is hard – too hard. The knowledge we seek is often fragmentary and disconnected, spread thinly across thousands of databases and millions of articles in thousands of journals. The intellectual energy required to search this array of data-archives, and the time and money this wastes, has led several researchers to challenge the methods by which we traditionally commit newly acquired facts and knowledge to the scientific record. We present some of these initiatives here – a whirlwind tour of recent projects to transform scholarly publishing paradigms, culminating in Utopia and the Semantic Biochemical Journal experiment. With their promises to provide new ways of interacting with the literature, and new and more powerful tools to access and extract the knowledge sequestered within it, we ask what advances they make and what obstacles to progress still exist? We explore these questions, and, as you read on, we invite you to engage in an experiment with us, a real-time test of a new technology to rescue data from the dormant pages of published documents. We ask you, please, to read the instructions carefully. The time has come; you may turn over your papers ...

Key words: dynamic document content, interactive PDF, linking documents with research data, manuscript mark-up, mark-up standards, semantic publishing.

INSTRUCTIONS TO READERS

Before reading any further, we are going to ask you to download a piece of software. Together, as we journey through this article, we will see the software is a new PDF document reader, called Utopia Documents (UD) in different scenarios. You are, of course, free to read on without installing the software; however, for those of you reading the PDF version of this article, seen through the lens of UD, much more functionality will be revealed and the text will become tantalizingly more interesting.

To install UD, please visit the abstract page for this article (at www.biochemj.org) or <http://www.biochemj.org/utopia>. The installation process is straightforward: simply follow the link to the website, and the guidance notes there will talk you through the software installation for your platform of choice.

Once you have successfully downloaded UD, you are ready to read on. As you do so, look out for the UD logo: this is used to draw your attention to interactive features, pinpointing where to click on particular icons. During the text, the story unfolds gradually and the interactive features will grow in complexity. We invite you to explore the increasing functionality at your leisure (for the more adventurous, full documentation is available from the installation site).

INTRODUCTION

New technologies that promise to transform our lives excite us, but often come with unanticipated side-effects. Just think about life before email, laptop computers or mobile phones and it's clear that as much as they've improved some aspects of our lives, they've made significant demands on us in others: e.g. to learn how to use yet another new gadget, to navigate yet another new interface, to cope with the daily bombardment of (often irrelevant) communications – in short, to control the technology before it controls us. Getting the balance right can be a struggle.

The life sciences have not been immune from these effects. Technological advances have led to the accumulation of data on a scale unthinkable only a couple of decades ago, promising to revolutionize how we 'do' biology and to have dramatic impacts on our understanding of such processes as gene expression, drug discovery, and the progression and treatment of disease [1,2]. Yet the megatons of data used to describe the phenomenal pace of data acquisition (from data floods [3], deluges [4,5], surging oceans [6] and tsunamis [7], to icebergs [8,9], avalanches [10], earthquakes [11] and explosions [12]) being a deep concern despite the early warnings, we appear to have been caught unprepared, and the resulting torrent of information has all but

Identifying this article
Resolved metadata for this article

"Calling International Rescue: knowledge lost in literature and data landslide!"

PID: [BJ20091474](#)
DOI: [10.1042/BJ20091474](#)
ISSN: [1470-8728](#)
Pubmed: [19929850](#)
PMC: [PMC2805925](#)

CrossRef
Formatted citation for this article

Attwood, Teresa K., Douglas B. Kell, Philip McDermott, James Marsh, Steve R. Pettifer, and David Thorne. Calling International Rescue: Knowledge Lost in Literature and Data Landslide! *Biochemical Journal* 424, no. 3 (December 15, 2009): 317–333. doi:10.1042/BJ20091474.

Altmetric
Who is talking about this article?

Mendeley
Related articles

Pritchard, L. (1989) *International rescue*. *Zeitschrift für ärztliche Fortbildung und Qualitätssicherung*. [\[Link\]](#)

Budde, Ulrich and Bergmann, Frauke and Michiels, Jan J. (2002) *Acquired von Willebrand syndrome: experience from 2 years in a single laboratory compared with data from the literature and an international registry*. *Seminars in Thrombosis And Hemostasis*. [\[Link\]](#)

Williams, AM. (2006) *Lost in translation? International migration, learning and knowledge*. *Progress in Human Geography*. [\[Link\]](#)

Miles, M F. (2001) *Microarrays: lost in a storm of data?* *Nature Reviews Neuroscience*. [\[Link\]](#)

Hoolahan, Jane. (2004) *A pivotal role in international rescue*. *Professional Nurse London England*. [\[Link\]](#)

[See more in Mendeley...](#)

SherpaRoMEO
Archiving status is 'White'.

This Biochemical Journal article, published by Portland Press, is classified as being **RoMEO white**. This means that archiving of this article not formally supported.

Other details and conditions apply.

Bibliography
From The Semantic Biochemical Journal

Andrade, M.; Sander, C. (1997) *Bioinformatics: from genome data to biological knowledge*. *Curr. Opin. Biotechnol.*, 8. [\[Find\]](#)

Antezana, E.; Kulper, M.; Mironov, V. (2009) *Biological knowledge management: the emerging role of the Semantic Web technologies*. *Brief. Bioinform.*, 10. [\[Find\]](#)

Ashburner, M.; Ball, C. A.; Blake, J. A.; Botstein, D.; Butler, H.; Cherry, J. M.; Davis, A. P.; Dolinski, K.; Dwight, S. S.; Eppig, J. T. (2000) *Gene ontology: tool for the unification of biology*. *The Gene Ontology Consortium. Nat. Genet.*, 25. [\[Find\]](#)

Asher, J. (1958) *Why are medical journals so dull?* *Br. Med. J.*, i. [\[Find\]](#)

Attwood, T. K. (2000) *The Babel of bioinformatics*. *Science*, 290. [\[Find\]](#)

Attwood, T. K.; Miller, C. J. (2002) *Progress in bioinformatics and the importance of being earnest*. *Bioessays*, 24. [\[Find\]](#)

318 T. K. Attwood and others

burst our databanks [1,3,14]. Desperate as things may seem, this is probably just a prelude to further troubles ahead, with 'week-long sequencing' becoming a reality, and the latest machines delivering terabytes of data per hour. Faced with this onslaught, standard laboratory information-management systems will be unable to cope, a situation that has been likened to "taking a drink from a fire hose" [5].

Beyond the information-management headaches [8] and nightmares [15], however, lies a deeper problem. Merely increasing the amounts of information we collect does not in itself become an increase in knowledge. For information to be usable, it must be stored and organized in ways that allow us to access it, to analyse it, to annotate it and to relate it to other information; only then can we begin to understand what it means; only with the acquisition of meaning do we acquire knowledge. The real problem is that we have failed to store and organize much of the rapidly accumulating information (whether in databases or documents) in rigorous, principled ways, so that finding what we want and understanding what's already known become exhausting, frustrating, stressful [7] and increasingly costly experiences.

Figure 1 Graphical illustration of the growth of biological research publications (incl. current total > 180 million), alongside the accumulation of research data, including nucleic acid sequences (black, current total ~180 million), computer-assisted protein sequences (magenta, current

Utopia Documents - bic005-09.dvi
317

Bj www.biochemj.org
UJ Central

Biochem J (2009) 424, 317-323 (Printed in Great Britain) doi: 10.1042/BJ20091474
317

REVIEW ARTICLE

Calling International Rescue: knowledge lost in literature and data landslide!

Teresa K. ATTWOOD*¹, Douglas B. KELLY, Philip McDERMOTT[†], James MARSH, Steve R. PETTIFER[†] and David THORNE[†]

¹School of Computer Science, The University of Manchester, Oxford Road, Manchester M13 9PL, UK, [†]Faculty of Life Sciences, The University of Manchester, Oxford Road, Manchester M13 9PL, UK

Discussion

Calling International Rescue: knowledge lost in literature and data landslide!

Mr Adam Marshall 134 weeks ago

This is a great paper for getting an overview of the capabilities of Utopia Documents, but remember that there have been additional functions added since this paper was published such as the comment function.

Reply

Simon Buckingham Shum 91 weeks ago

(test comment) Utopia looks like a significant step forward for scientific publishing. Important to integrate new tech with old practices

Reply

Miss Judy Chen 64 weeks ago

does my comment get saved as part of the thread?

Semantic search provides and more requested features to you read a real-time at pages of interactive papers...

IF linking & mark-up

specific us, think about us and it's if our lives, g. to learn not for new (irrelevant) before it

the effects, of data on

Identifying this article

Resolved metadata for this article

"Calling International Rescue: knowledge lost in literature and data landslide!"

PII: BJ20091474
DOI: 10.1042/BJ20091474
ISSN: 1470-8728
Pubmed: 19929850
PMC: PMC2805925

CrossRef

Formatted citation for this article

Atwood, Teresa K., Douglas B. Kelly, Philip McDermott, James Marsh, Steve R. Pettifer, and David Thorne. Calling International Rescue: Knowledge Lost in Literature and Data Landslide. *Biochemical Journal* 424, no. 3 (December 15, 2009): 317-333. doi:10.1042/BJ20091474.

chicago-fullnote-bibliography

Altmetric

Who is talking about this article?

Mendeley

Related articles

Pritchard, L. (1989) *International rescue*. Zeitschrift für ärztliche Fortbildung und Qualitätssicherung. [Link]

Budde, Ulrich and Bergmann, Frauke and Michels, Jan J. (2002) *Acquired von Willebrand syndrome: experience from 2 years in a single laboratory compared with data from the literature and an international registry*. Seminars in Thrombosis And Hemostasis. [Link]

Williams, AM. (2006) *Lost in translation? International migration, learning and knowledge*. Progress in Human Geography. [Link]

Miles, M F. (2001) *Microarrays: lost in a storm of data?* Nature Reviews Neuroscience. [Link]

Hoolahan, Jane. (2004) *A pivotal role in international rescue*. Professional Nurse London England. [Link]

See more in Mendeley...

Sherpa/RoMEO

Archiving status is 'white'

This Biochemical Journal article, published by Portland Press, is classified as being RoMEO white. This means that archiving of this article not formally supported.

Other details and conditions apply.

Bibliography

From The Semantic Biochemical Journal

Andrade, M.; Sander, C., (1997) *Bioinformatics: from genome data to biological knowledge*. Curr. Opin. Biotechnol., 8. [Find]

Antezana, E.; Kuiper, M.; Mironov, V., (2009) *Biological knowledge management: the emerging role of the Semantic Web technologies*. Brief. Bioinform., 10. [Find]

Ashburner, M.; Ball, C. A.; Blake, J. A.; Botstein, D.; Butler, H.; Cherry, J. M.; Davis, A. P.; Dolinski, K.; Dwight, S. S.; Eppig, J. T., (2000) *Gene ontology: tool for the unification of biology*. The Gene Ontology Consortium. Nat. Genet., 25. [Find]

Asher, J., (1958) *Why are medical journals so dull?* Br. Med. J., 8. [Find]

Atwood, T. K., (2000) *The Babel of bioinformatics*. Science, 290. [Find]

Atwood, T. K.; Miller, C. J., (2002) *Progress in bioinformatics and the importance of being earnest*.

318 T. K. Atwood and others

hurst our databanks [1,3,4]. Desperate as things may seem, this is probably just a prelude to further troubles ahead, with 'desk-top sequencing' becoming a reality, and the latest machines delivering terabytes of data per hour. Faced with this onslaught, standard laboratory information-management systems will be unable to cope, a situation that has been likened to "taking a drink from a fire hose" [5].

Beyond the information-management headaches [8] and nightmares [15], however, lies a deeper problem. Merely increasing the amounts of information we collect does not in itself bestow an increase in knowledge. For information to be usable, it must be stored and organized in ways that allow us to access it, to analyse it, to annotate it and to relate it to other information; only then can we begin to understand what it means, only with the acquisition of meaning do we acquire knowledge. The real problem is that we have failed to store and organize much of the rapidly accumulating information (whether in databases or documents) in rigorous, principled ways, so that finding what we want and understanding what's already known become exhausting, frustrating, stressful [7] and increasingly costly experiences.

Figure 1 Graphical illustration of the growth of biological research publications (red, current total ~19 million), alongside the accumulation of research data, including nucleic acid sequences (black, current total ~163 million), computer-annotated protein sequences (magenta, current

The winners of the contest developed a tool (Utopia) that addresses the routine need of life scientists to be able both to jump from gene or protein names to their molecular sequences, and to understand more about particular genes, proteins or small molecules encountered in the literature [14]. With a single mouse click, Reflect tags such entities when they occur in webpages; it does this by drawing on a large, consolidated dictionary

A rather different slant on the problem of dissemination and reuse of scientific knowledge is offered by the Liquid Publication Project, a European initiative paraded by Springer Verlag [10]. The intention here is for publications to become fluid entities, created in a collaborative and evolutionary fashion over time, in much the same way as open-source software is developed.

© 2008 The Author(s)
This article is published with the permission of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/3.0/). All rights reserved. No commercial use, distribution and reproduction in any medium, provided the original work is properly cited.

Resolving knowledge lost in literature and data 325

Identifying this article
Resolved metadata for this article

"Calling International Rescue: knowledge lost in literature and data landslide!"
PII: BJ20091474
DOI: 10.1042/BJL20091474
ISSN: 1470-8728
Pubmed: 19929650
PMC: PMC2605925

CrossRef
Formatted citation for this article

Athwood, Teresa K., Douglas B. Kell, Philip McDermott, James Marsh, Steve R. Pettifer, and David Thorne. Calling International Rescue: Knowledge Lost in Literature and Data Landslide! *Biochemical Journal* 424, no. 3 (December 15, 2009): 317-333. doi:10.1042/BJL20091474.

Altmetric
Who is talking about this article?

Mendeley
Related articles

Pritchard, L. (1989) *International rescue. Zeitschrift für ärztliche Fortbildung und Qualitätssicherung*. [Link]

Budde, Ulrich and Bergmann, Frauke and Michiels, Jan J. (2002) *Acquired von Willebrand syndrome: experience from 2 years in a single laboratory compared with data from the literature and an international registry. Seminars in Thrombosis And Hemostasis*. [Link]

Williams, AM. (2008) *Lost in translation? International migration, learning and knowledge. Progress in Human Geography*. [Link]

Miles, M. F. (2001) *Microarrays: lost in a storm of data? Nature Reviews Neuroscience*. [Link]

Hoolahan, Jane. (2004) *A pivotal role in international rescue. Progress in Human Geography*. [Link]

Sherpa/ReMEO
Archiving status is 'white'.
This Biochemical Journal article, published by Portland Press, is classified as being ReMEO white. This means that archiving of this article is not formally supported.
Other details and conditions apply.

Bibliography
From The Semantic Biochemical Journal

Andrade, M.; Sander, C. (1997) *Bioinformatics: from genome data to biological knowledge. Curr. Opin. Biotechnol.*, 8. [FIND]

Antezana, E.; Kulper, M.; Mironov, V. (2009) *Biological knowledge management: the emerging role of the Semantic Web technologies. Brief. Bioinform.*, 10. [FIND]

Ashburner, M.; Ball, C. A.; Blake, J. A.; Botstein, D.; Butler, H.; Cherry, J. M.; Davis, A. P.; Dolinski, K.; Dwight, S. S.; Eppig, J. T. (2000) *Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat. Genet.*, 25. [FIND]

Asher, J. (1958) *Why are medical journals so dull? Br. Med. J.*, 1. [FIND]

Athwood, T. K. (2000) *The Babel of bioinformatics. Science*, 290. [FIND]

Athwood, T. K.; Miller, C. J. (2002) *Progress in bioinformatics and the importance of being earnest. Biochem. Soc. Rev.*, 6. [FIND]

© 2008 The Author(s)
This article is published with the permission of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/3.0/). All rights reserved. No commercial use, distribution and reproduction in any medium, provided the original work is properly cited.

326 T. K. Athwood and others

The winners of the contest developed a tool (Utopia) that addresses the routine need of life scientists to be able both to jump from gene or protein names to their molecular sequences, and to understand more about particular genes, proteins or small molecules encountered in the literature [14]. With a single mouse click, Reflect tags such entities when they occur in webpages; it does this by drawing on a large, consolidated dictionary

A rather different slant on the problem of dissemination and reuse of scientific knowledge is offered by the Liquid Publication Project, a European initiative paraded by Springer Verlag [10]. The intention here is for publications to become fluid entities, created in a collaborative and evolutionary fashion over time, in much the same way as open-source software is developed.

© 2008 The Author(s)
This article is published with the permission of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/3.0/). All rights reserved. No commercial use, distribution and reproduction in any medium, provided the original work is properly cited.

Resolving knowledge lost in literature and data 325

Identifying this article
Resolved metadata for this article

"Calling International Rescue: knowledge lost in literature and data landslide!"
PII: BJ20091474
DOI: 10.1042/BJL20091474
ISSN: 1470-8728
Pubmed: 19929650
PMC: PMC2605925

CrossRef
Formatted citation for this article

Athwood, Teresa K., Douglas B. Kell, Philip McDermott, James Marsh, Steve R. Pettifer, and David Thorne. Calling International Rescue: Knowledge Lost in Literature and Data Landslide! *Biochemical Journal* 424, no. 3 (December 15, 2009): 317-333. doi:10.1042/BJL20091474.

Altmetric
Who is talking about this article?

Mendeley
Related articles

Pritchard, L. (1989) *International rescue. Zeitschrift für ärztliche Fortbildung und Qualitätssicherung*. [Link]

Budde, Ulrich and Bergmann, Frauke and Michiels, Jan J. (2002) *Acquired von Willebrand syndrome: experience from 2 years in a single laboratory compared with data from the literature and an international registry. Seminars in Thrombosis And Hemostasis*. [Link]

Figure 1 Link images being able to tag between a published table of numerical values and their graphical representation

For notes, view this article online. For this special issue of the British Journal of Pharmacology, please see the result of a search on PubMed, S. M. M. et al. (2008) *International rescue: experience from 2 years in a single laboratory compared with data from the literature and an international registry. Seminars in Thrombosis And Hemostasis*. [Link]

Compound	A.M.Papp	RK	Fa%	Caco-2b-Papp	PAMPA-Papp	LogKo/w	LogD
Chlorothiazide	0.86±0.04	99.1	13	0.015	0.13	-0.24	-0.05
Acidovoril	0.91±0.02	99.9	21	0.025	0.00	-1.74	-1.86
Nadolol	1.37±0.03	99.5	32	0.038	0.00	0.71	0.68
Alpha-Methyl-dopa	0.32±0.01	97.1	41	0.015	0.00	-1.80	-1.80
Ranitidine	2.15±0.03	99.9	55	0.049	0.05	0.27	-0.29
Metformin	2.27±0.20	97.3	55	0.550	0.02	-1.43	-1.22
Furosemide	2.75±0.02	97.0	60	0.012	0.06	2.29	-0.69
Hydrochlorothiazide	3.10±0.05	98.3	70	0.051	0.00	-0.07	-0.12
Chloramphenicol	3.97±0.04	99.3	90	0.206	0.17	1.14	1.14
Hydrocortisone	4.28±0.07	99.8	91	1.40	0.34	1.61	1.55
Flindolol	3.74±0.07	99.2	92	1.67	0.49	1.75	0.19
Propionolol	3.97±0.08	99.8	93	4.19	2.35	1.25	1.25
Metoprolol	4.81±0.08	99.6	95	2.37	0.35	1.88	-0.16
Theophylline	4.05±0.06	99.1	97	2.52	0.48	-0.25	-0.05
Tripropthorin	4.55±0.09	99.8	98	3.30	0.50	0.91	0.74
Naxopren	4.88±0.02	98.9	97	8.90	1.06	3.18	0.23
Verapamil	4.16±0.03	97.5	98	1.58	0.74	3.79	2.66

Export as CSV... Toggle Table/Graph

Asher, J. (1958) *Why are medical journals so dull? Br. Med. J.*, 1. [FIND]

Athwood, T. K. (2000) *The Babel of bioinformatics. Science*, 290. [FIND]

Athwood, T. K.; Miller, C. J. (2002) *Progress in bioinformatics and the importance of being earnest. Biochem. Soc. Rev.*, 6. [FIND]

The author(s) has paid for this article to be freely available under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/2.5/>) which permits unrestricted non-commercial use, distribution and reproduction in any medium, provided the original work is properly cited.

326 T. K. Attwood and others

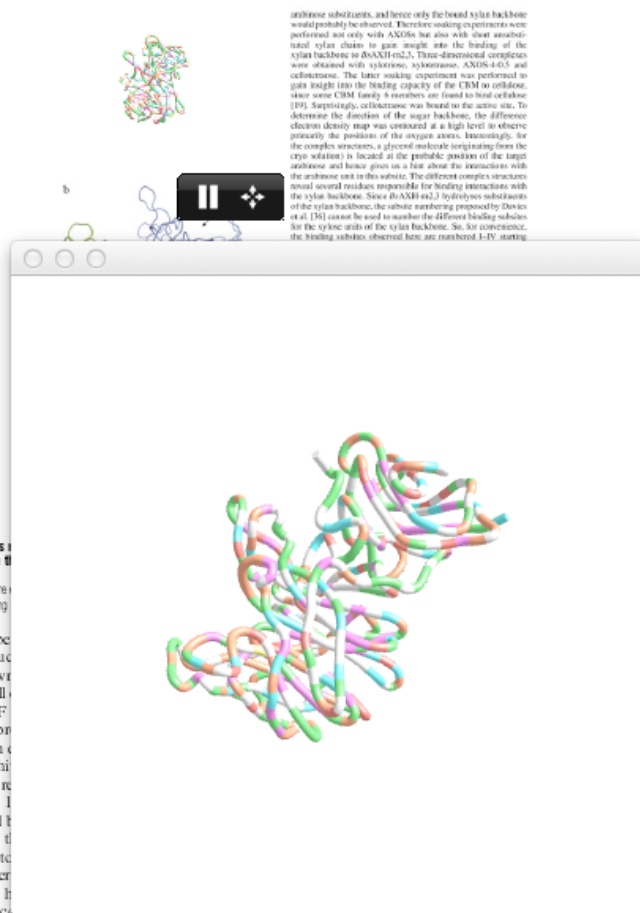


Figure 10 Bourne imagines the interactions described in the paper.

In this 2009 *BJ* paper, Vandermarck et al. describe a protein module in green. For readers viewing

why publications have been offered by such infrastructure the community has grown electronic articles are still limited forms, like PDF because of impeding the process.

To gain the most from dormant document archives clearly necessary. But legacy data is complex, it must therefore be found possible to introduce to the change it is reasonable to has not hitherto considered published articles, and providing the semantic context.

long-run, it is to be hoped that the benefits of semantic mark-up, and the availability of the right tools, will together help to seed this much-needed cultural change: compare and contrast, for example,

arabinose substrates, and hence only the bound sialin backbone would probably be observed. Therefore soaking experiments were performed not only with AXD56 but also with short acetylated sialin chains to gain insight into the binding of the sialin backbone to BAXH162A. Three-dimensional complexes were obtained with sialosyl, sialosylamine, AXD5-4-0.5 and cellobiose. The latter soaking experiment was performed to gain insight into the binding capacity of the CRM to cellobiose, since some CRM family 8 members are found to bind cellobiose [9]. Superimposing cellobiose was bound to the active site. To determine the direction of the sugar backbone, the difference electron density map was contoured at a high level to observe precisely the positions of the oxygen atoms. Incidentally, for the complex structures, a glycerol molecule originating from the cryo solution is located at the probable position of the target substrate and hence gives us a hint about the interactions with the arabinose unit in this substrate. The different complex structures reveal several residues responsible for binding interactions with the sialin backbone. Since BAXH162J hydrolyses substrates of the sialin backbone, the substrate numbering proposed by Davies et al. [10] cannot be used to number the different binding subsites for the sialic units of the sialin backbone. So, for convenience, the binding subsites observed here are numbered I-IV, starting

to explore

hydrate-binding

ologies for the same time especially a review of not at all relating and imaging the 'facts' may knowledge may however, we literature or re the rates papers, and ologies that asisencies, record the evolution of the current state of our knowledge, are therefore potentially profound. Consider, for a moment, the example illustrated in Figure 13.

Altmetric

Who is talking about this article?



Mendeley

Related articles



Pritchard, L. (1989) *International rescue*. Zeitschrift für ärztliche Fortbildung und Qualitätssicherung. [\[Link\]](#)

Budde, Ulrich and Bergmann, Frauke and Michiels, Jan J. (2002) *Acquired von Willebrand syndrome: experience from 2 years in a single laboratory compared with data from the literature and an international registry*. Seminars In Thrombosis And Hemostasis. [\[Link\]](#)

Williams, AM. (2006) *Lost in translation? International migration, learning and knowledge*. Progress in Human Geography. [\[Link\]](#)

Miles, M F. (2001) *Microarrays: lost in a storm of data?* Nature Reviews Neuroscience. [\[Link\]](#)

Hoolahan, Jane. (2004) *A pivotal role in international rescue*. Professional Nurse London England. [\[Link\]](#)

[See more in Mendeley...](#)

Sherpa/RoMEO

Archiving status is 'white'.



This Biochemical Journal article, published by Portland Press, is classified as being RoMEO white. This means that archiving of this article not formally supported.

Other [details and conditions](#) apply.

Bibliography

From The Semantic Biochemical Journal



Andrade, M.; Sander, C.. (1997) *Bioinformatics: from genome data to biological knowledge*. Curr. Opin. Biotechnol., 8. [\[Find\]](#)

Antezana, E.; Kuiper, M.; Mironov, V.. (2009) *Biological knowledge management: the emerging role of the Semantic Web technologies*. Brief. Bioinform., 10. [\[Find\]](#)

Ashburner, M.; Ball, C. A.; Blake, J. A.; Botstein, D.; Butler, H.; Cherry, J. M.; Davis, A. P.; Dolinski, K.; Dwight, S. S.; Eppig, J. T.. (2000) *Gene ontology: tool for the unification of biology*. The Gene Ontology Consortium. Nat. Genet., 25. [\[Find\]](#)

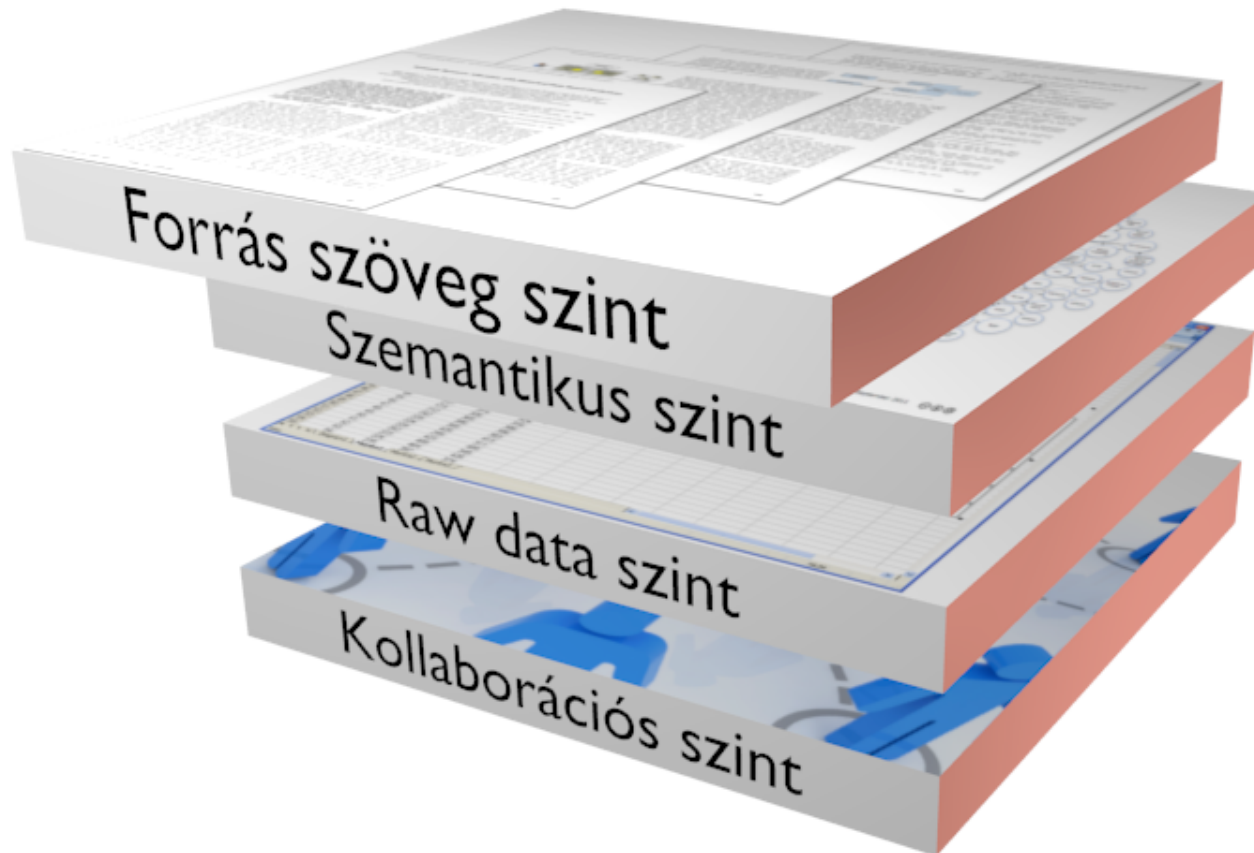
Asher, J.. (1958) *Why are medical journals so dull?* Br.

- Research Object (RO)
 - eScience csomagolás (konténer) formátum
 - Reusable, Repurposable, Repeatable (with different datasets), Reproducible, Replay-able (in silicon)
 - Research Objects for Scholarly Communication (ROSC) W3C activity, FORCE11, researchobject.org

NextGeneration

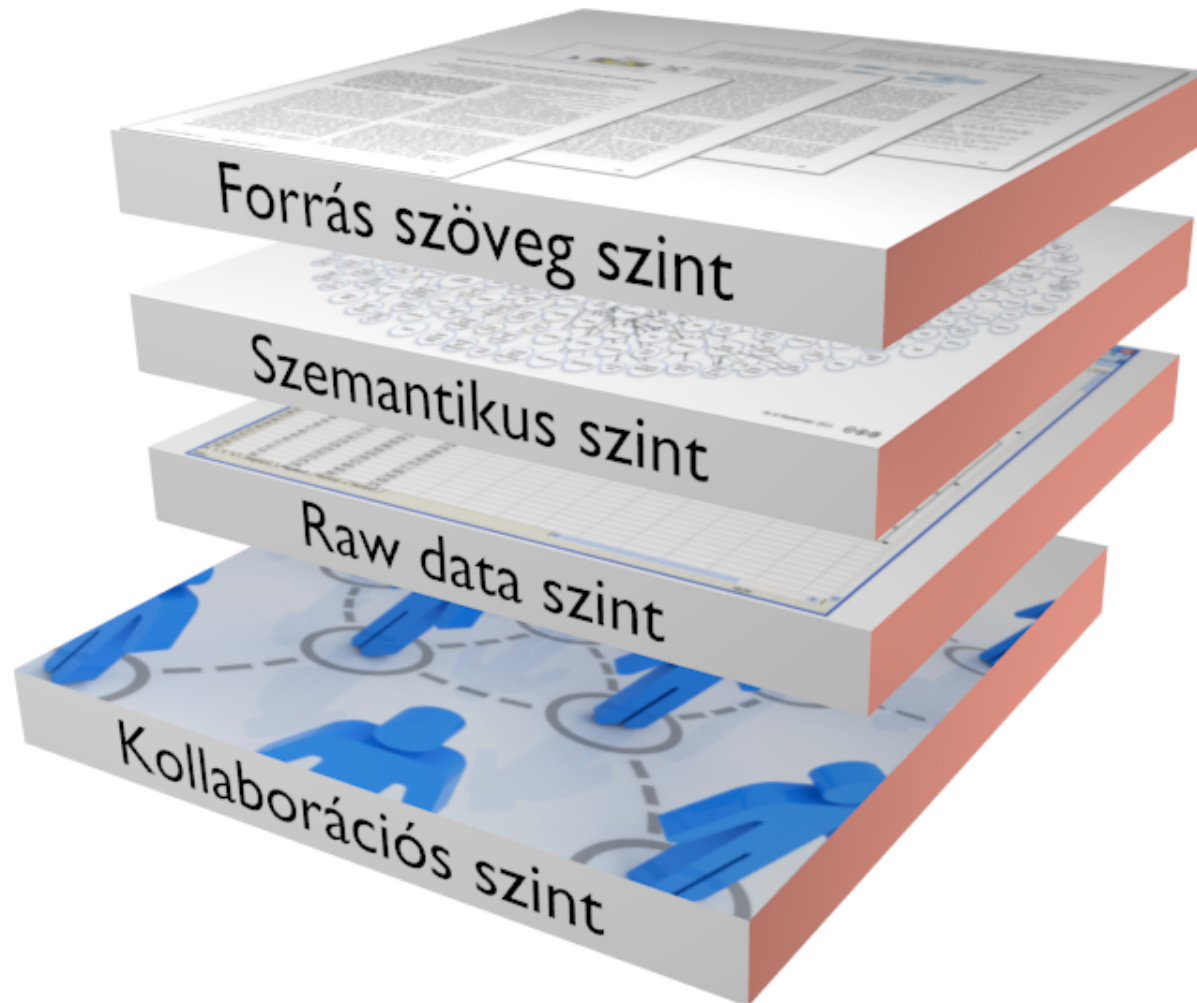
Interaktív publikációkat (RO) kezelni
képes tudományos folyóirat
infrastruktúra

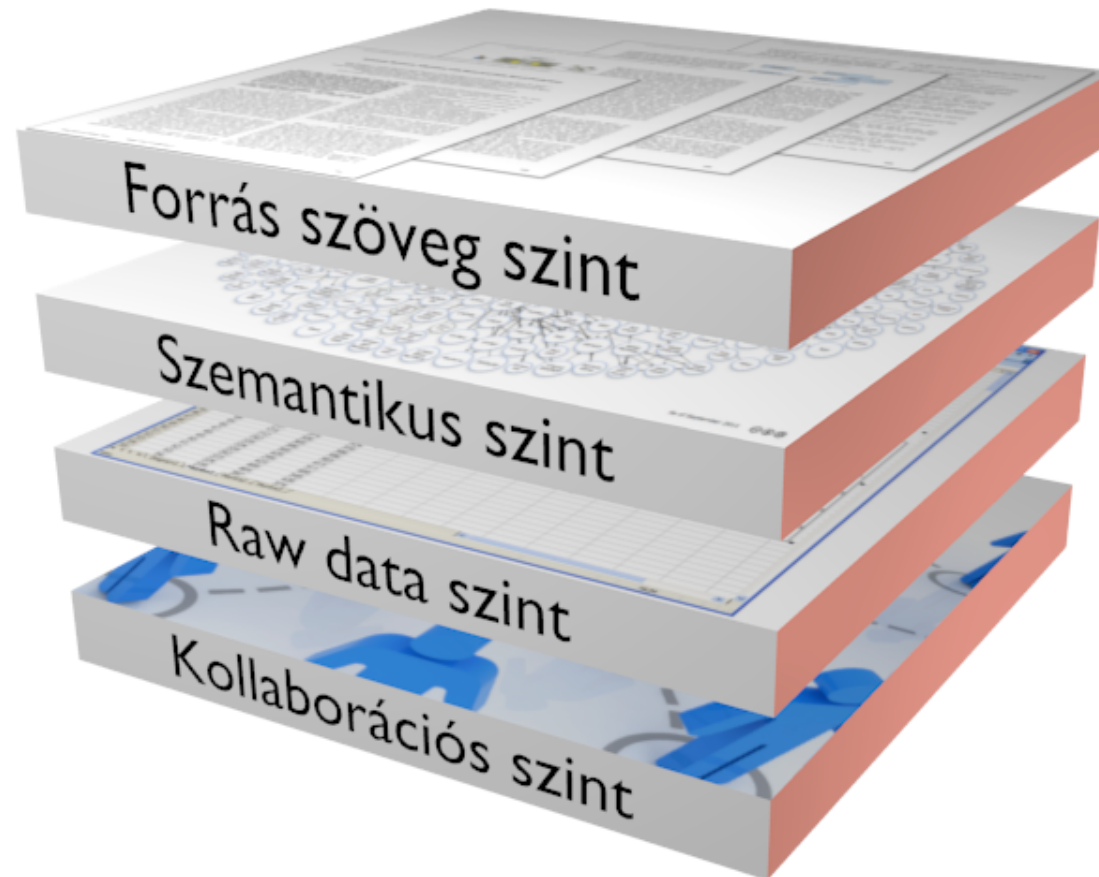


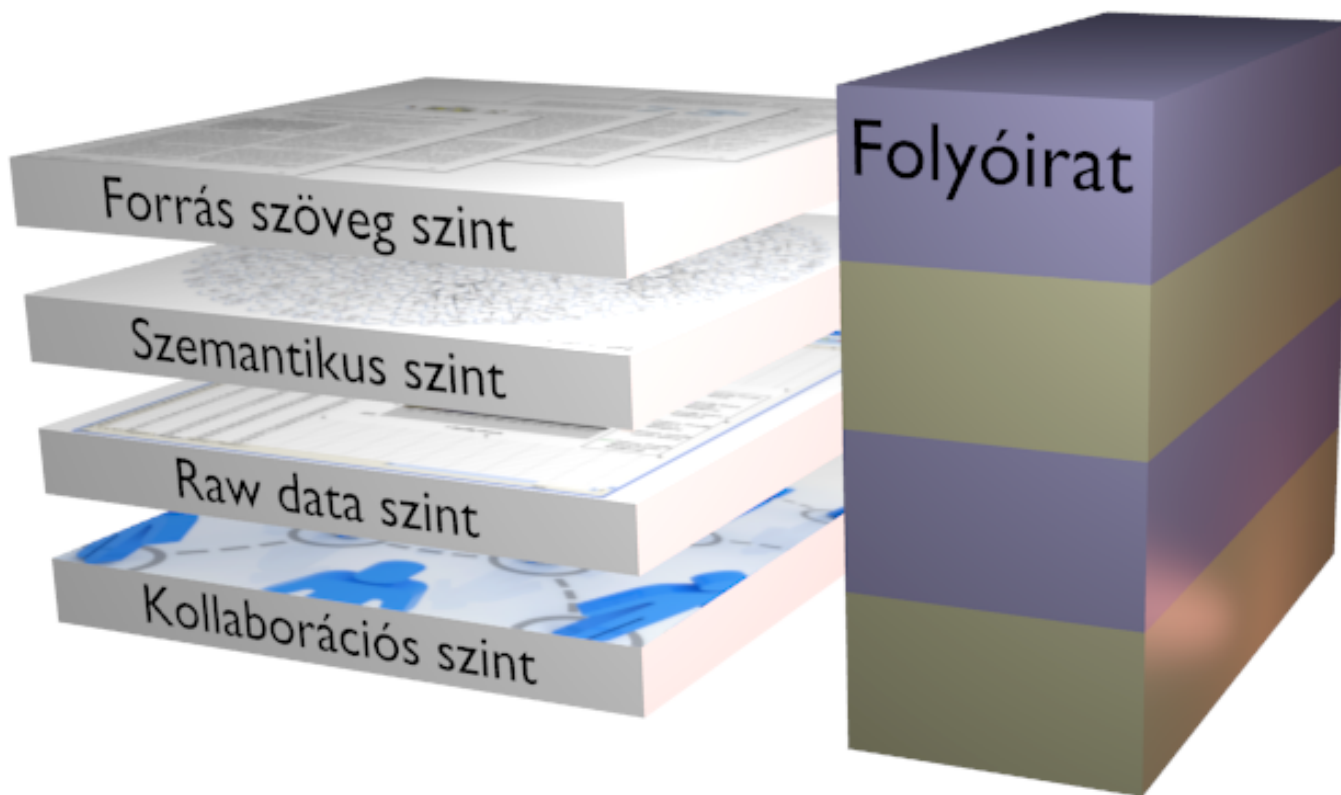


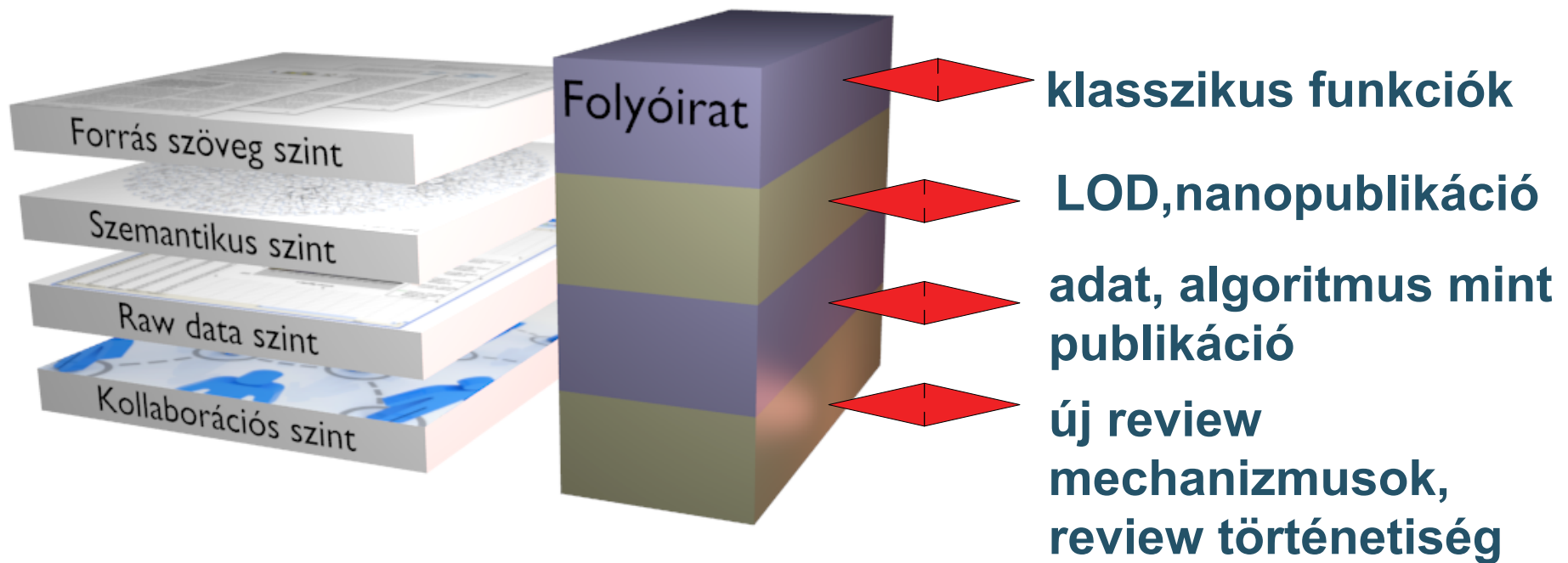












ELEK PÉTER–KISS ÁRON

Az ingatlanadó a magyar adórendszerben

Elméleti megfontolások és mikroszimulációs elemzés

A tanulmány az ingatlanadó hatásait elemzi a magyar adórendszer keretein belül. Vizsgálja a 2009-ben elfogadott (és azóta az Alkotmánybíróság által részben megsemmisített) vagyonadó szerinti változatot, valamint hipotetikus alternatív változatokat. Nemzetközi összehasonlításban is bemutatja a Magyarországon jelenleg hatályos vagyont terhelő elvonásokat, majd összefoglalja az adóteher-megoszlás (adóincidencia) közgazdaságtanának ingatlanadóra vonatkozó megállapításait. Az empirikus elemzés a KSH háztartási költségvetési felvételeinek (HKF) adatain szimulálja az adófizetési teher jövedelmi és más demográfiai csoportok szerinti eloszlását. A tanulmány végül összefoglalja az ingatlanadó mellett és ellen felhozható közgazdasági érveket.*

Journal of Economic Literature (JEL) kód: H20, H22.

A tanulmány célja, hogy az ingatlanadó hatásait elemezze a magyar adórendszer keretein belül. Az elmúlt években több ingatlanadó-koncepció jutott el a parlamenti döntéshozásig, legutoljára a 2009-ben elfogadott, 2010. január elsején hatályba lépett, de 2010 januárjában az Alkotmánybíróság által részben megsemmisített vagyonadó részeként. Tanulmányunkban egyaránt elemezzük az ingatlanadót „általában” és konkrétan a 2009-ben elfogadott változatot.

Korábbi hazai tanulmányokhoz képest (*Kopint-Tárki* [2007], *Szalai* [2008]) több figyelmet fordítunk az ingatlanadó gazdasági terhének elméleti és adófizetési terhének empirikus elemzésére, és kevesebbet az ingatlanadóval kapcsolatban felmerülő számos adótechnikai kérdésre.

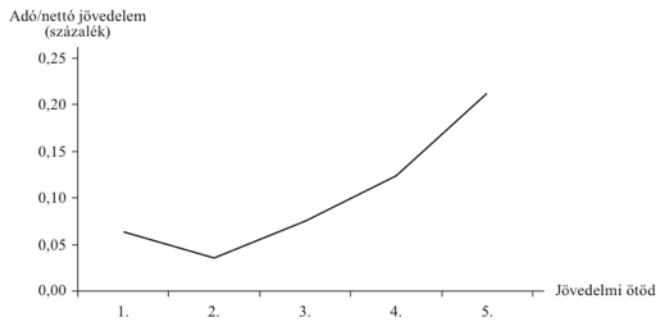
Először bemutatjuk, valamint nemzetközi összehasonlításban elhelyezzük a Magyarországon jelenleg hatályos vagyont terhelő elvonásokat. Az elemzésből kiderül, hogy Magyarországon – hasonlóan a visegrádi országokhoz – a munkát és a fogyasztást terhelő elvonások adóbevételeken belüli súlya az EU-tagországok átlagához képest nagyobb, a tőkét terhelőké viszont kisebb. Az ingatlanadón terheken belül Magyarország nagyobb mértékben támaszkodik a forgalmi jellegű illetékekre, mint más OECD-tagállamok.

Külön fejezetben foglaljuk össze az adóteher-megoszlás (adóincidencia) ingatlanadóra vonatkozó közgazdaságtani eredményeit, majd egy másikban ismertetjük a kutatás empirikus részét, amely a KSH háztartási költségvetési felmérésének (HKF) adataival szimulálja az adófizetési teher jövedelmi és egyéb demográfiai csoportok szerinti eloszlását – a

* A szerzők szeretnének köszönetet nyilvánítani *Szalai Ákosnak*, *Kiss Gergelynek* és *Scharle Ágotának* a tanulmány egy korábbi változatához fűzött megjegyzéseikért, *Kanyó Lórántnak* és *Nagy Imrénének* adóstatistikai észrevételeikért és a tanulmány műhelyvitáján megjelenteknek észrevételeikért. A szerzők a kutatás idején a Pénzügyminisztérium munkatársai voltak. A tanulmány a szerzők véleményét tükrözi.

4. ábra

A háztartások jövedelemarányos terhelése a hatályba lépett és megsemmisített vagyonadó szerint



Forrás: saját számítás a háztartási költségvetési felvétel alapján.

Ezek után vizsgáljuk meg néhány hipotetikus ingatlanadó-változat újraelosztási hatását, egyelőre csak a lakásokra összpontosítva. Az A) változat egy egykulcsos, 0,2 százalékos lakásadó, a B) és C) változatban a lakások progresszív kulcs szerint adóznak a 3. táblázat

Sáv (millió forint)

Változat	0-10	10-20	20-30	30-50	50-
A	0,2	0,2	0,2	0,2	0,2
B	0,15	0,3	0,5	0,5	1
C	0	0,6	0,6	1	1

háztartásokra kifejtett aggregált hatásait. Ugyanakkor az egyes változatok a különböző jövedelmi helyzetű csoportokra lényegesen eltérő hatást gyakorolnak. A lakásvagyon eloszlásából természetesen következik, hogy a magasabb jövedelmű háztartások összességében többet fizetnek mindegyik változatban, mint az alacsonyabb jövedelműek. A jövedelemarányos terhelést az 5. ábra mutatja: az A) változat mindkét [törvény szerinti (az ábra a) része), illetve regressziós modellből adódó (az ábra b) része)] értékszámítási mód szerint regresszív, a B) változat közelítően jövedelemarányos, a C) pedig erősen progresszív hatású.

Más csoportosításban vizsgálva a lakásadó hatását, az egykulcsos változat a különböző településtípusok háztartásait jövedelemarányosan hasonlóan érinti, míg a többkulcsos esetek egyértelműen a budapesti háztartásokat érintik hátrányosabban (6. ábra).

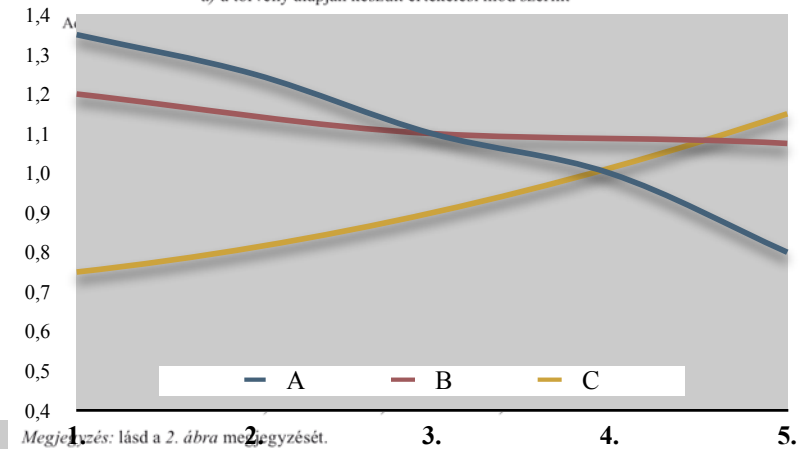
Ha az adófizetés eloszlását a gyermekszám és gazdasági aktivitás szerint vizsgáljuk – az egyszerűség kedvéért csak az egykulcsos A) és a leginkább progresszív C) változatra összpontosítva –, akkor az inaktív háztartásokat¹⁵ az egykulcsos adó esetén jövedelemarányosan

¹⁵ Azt a háztartást nevezzük aktívnak, ahol van foglalkoztatott tag, míg az összes többi háztartás inaktív – ez utóbbiak nagy része nyugdíjas.

5. ábra

A háztartások jövedelemarányos terhelése a három változatban

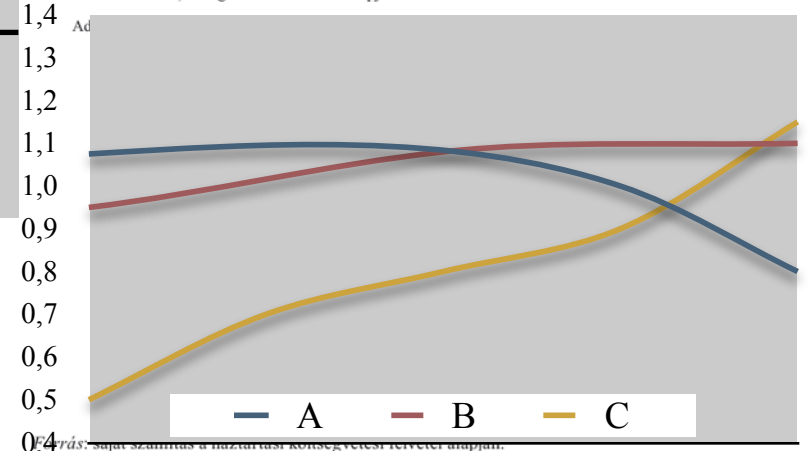
a) a törvény alapján készült értékelési mód szerint



Megjegyzés: lásd a 2. ábra megjegyzését.

Forrás: saját számítás a 2005. évi háztartási költségvetési felvétel alapján.

b) a regressziós modell alapján készült értékelési mód szerint



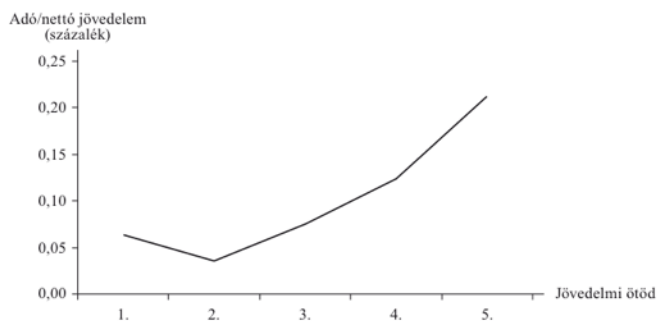
Forrás: saját számítás a háztartási költségvetési felvétel alapján.

nagyobb terhelés éri, mint az aktív háztartásokat, ugyanakkor ez a különbség a leginkább progresszív esetben összességében eltűnik. A gyermekszám¹⁶ növekedése összességében csökkenti a jövedelemarányos terhelést az A) változatban, de nincs nagy jelentősége a C) változat szerint.

¹⁶ A háztartásban lévő 20 év alatti eltartott gyerekek száma.

4. ábra

A háztartások jövedelemarányos terhelése a hatályba lépett és megsemmisített vagyonadó szerint



Forrás: saját számítás a háztartási költségvetési felvétel alapján.

Ezek után vizsgáljuk meg néhány hipotetikus ingatlanadó-változat újraelosztási hatását, egyelőre csak a lakásokra összpontosítva. Az A) változat egy egykulcsos, 0,2 százalékos lakásadó, a B) és C) változatban a lakások progresszív kulcs szerint adóznak a 3. táblázat

Sáv (millió forint)

Változat	0-10	10-20	20-30	30-50	50-
A	0,2	0,2	0,2	0,2	0,2
B	0,15	0,3	0,5	0,9	1
C	0	0,6	0,6	1	1

háztartásokra kifejtett aggregált hatásait hasonlító. Ugyanakkor az egyes változatok a különböző jövedelmi helyzetű csoportokra lényegesen eltérő hatást gyakorolnak. A lakásvagyon eloszlásából természetesen következik, hogy a magasabb jövedelmű háztartások összességében többet fizetnek mindegyik változatban, mint az alacsonyabb jövedelműek. A jövedelemarányos terhelést az 5. ábra mutatja: az A) változat mindkét [törvény szerinti (az ábra a) része), illetve regressziós modellből adódó (az ábra b) része)] értékszámítási mód szerint regresszív, a B) változat közelítően jövedelemarányos, a C) pedig erősen progresszív hatású.

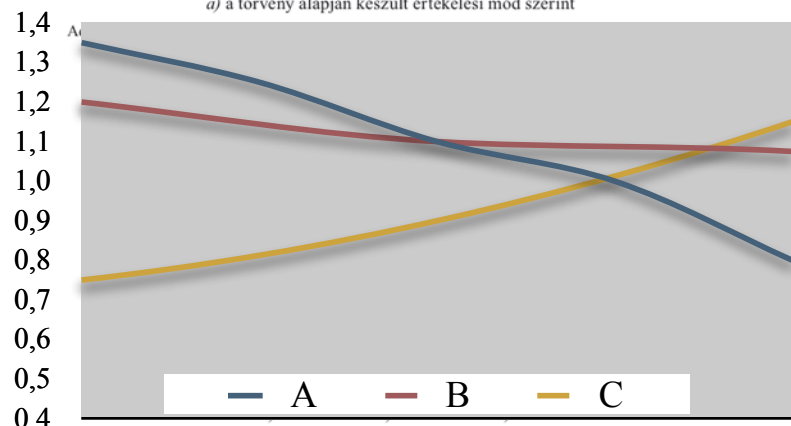
Más csoportosításban vizsgálva a lakásadó hatását, az egykulcsos változat a különböző településtípusok háztartásait jövedelemarányosan hasonlóan érinti, míg a többkulcsos esetek egyértelműen a budapesti háztartásokat érintik hátrányosabban (6. ábra).

Ha az adófizetés eloszlását a gyermekszám és gazdasági aktivitás szerint vizsgáljuk – az egyszerűség kedvéért csak az egykulcsos A) és a leginkább progresszív C) változatra összpontosítva –, akkor az inaktív háztartásokat¹⁵ az egykulcsos adó esetén jövedelemarányosan

¹⁵ Azt a háztartást nevezzük aktívnak, ahol van foglalkoztatott tag, míg az összes többi háztartás inaktív – ez utóbbiak nagy része nyugdíjas.

5. ábra

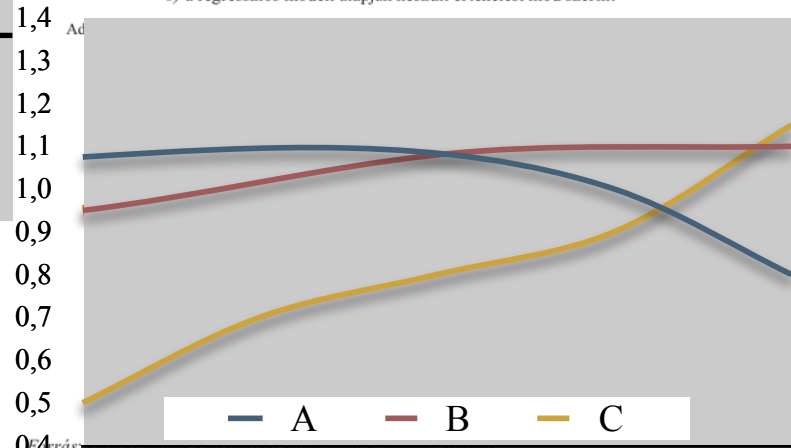
A háztartások jövedelemarányos terhelése a három változatban
a) a törvény alapján készült értékelési mód szerint



Megjegyzés: lásd a 2. ábra megjegyzését.

Forrás: saját számítás a 2005. évi háztartási költségvetési felvétel alapján.

b) a regressziós modell alapján készült értékelési mód szerint



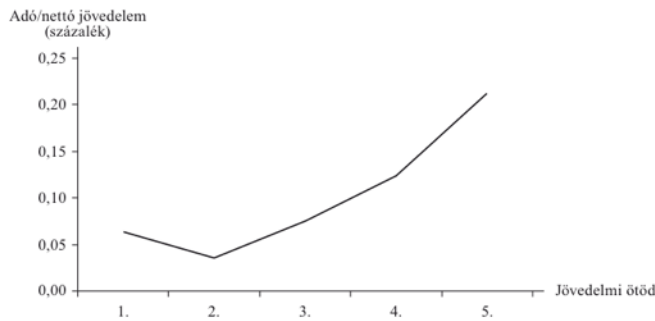
Forrás: saját számítás a háztartási költségvetési felvétel alapján.

nagyobb terhelés éri, mint az aktív háztartásokat, ugyanakkor ez a különbség a leginkább progresszív esetben összességében eltűnik. A gyermekszám¹⁶ növekedése összességében csökkenti a jövedelemarányos terhelést az A) változatban, de nincs nagy jelentősége a C) változat szerint.

¹⁶ A háztartásban lévő 20 év alatti eltartott gyerekek száma.

4. ábra

A háztartások jövedelemarányos terhelése a hatályba lépett és megsemmisített vagyonadó szerint



Forrás: saját számítás a háztartási költségvetési felvétel alapján.

Ezek után vizsgáljuk meg néhány hipotetikus ingatlanadó-változat újraelosztási hatását, egyelőre csak a lakásokra összpontosítva. Az A) változat egy egykulcsos, 0,2 százalékos lakásadó, a B) és C) változatban a lakások progresszív kulcs szerint adóznak a 3. táblázat

Változa	Sáv (millió forint)				
t	0-10	10-20	20-30	30-50	50-
A	0,2	0,2	0,2	0,2	0,2
B	0,15	0,3	0,5	0,9	1,5
C	0	0,6	0,6	1	1

háztartásokra kifejtett aggregált hatásait hasonlító. Ugyanakkor az egyes változatok a különböző jövedelmi helyzetű csoportokra lényegesen eltérő hatást gyakorolnak. A lakásvagyon eloszlásából természetesen következik, hogy a magasabb jövedelmű háztartások összességében többet fizetnek mindegyik változatban, mint az alacsonyabb jövedelműek. A jövedelemarányos terhelést az 5. ábra mutatja: az A) változat mindkét [törvény szerinti (az ábra a) része), illetve regressziós modellből adódó (az ábra b) része)] értékszámítási mód szerint regresszív, a B) változat közelítően jövedelemarányos, a C) pedig erősen progresszív hatású.

Más csoportosításban vizsgálva a lakásadó hatását, az egykulcsos változat a különböző településtípusok háztartásait jövedelemarányosan hasonlóan érinti, míg a többkulcsos esetek egyértelműen a budapesti háztartásokat érintik hátrányosabban (6. ábra).

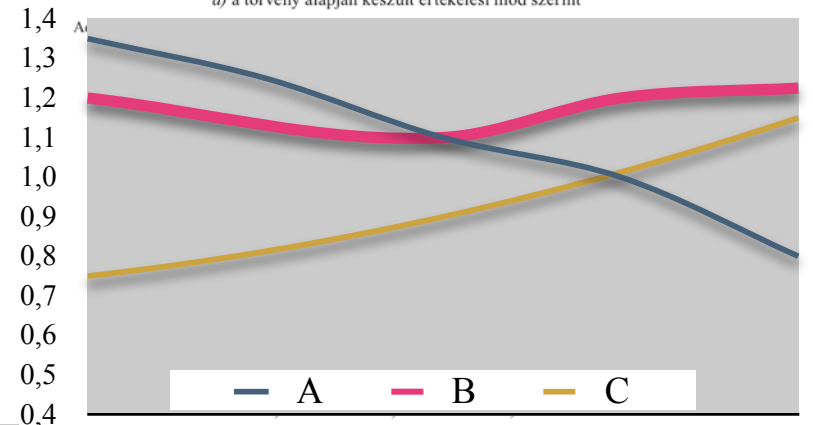
Ha az adófizetés eloszlását a gyermekszám és gazdasági aktivitás szerint vizsgáljuk – az egyszerűség kedvéért csak az egykulcsos A) és a leginkább progresszív C) változatra összpontosítva –, akkor az inaktív háztartásokat¹⁵ az egykulcsos adó esetén jövedelemarányosan

¹⁵ Azt a háztartást nevezzük aktívnak, ahol van foglalkoztatott tag, míg az összes többi háztartás inaktív – ez utóbbiak nagy része nyugdíjas.

5. ábra

A háztartások jövedelemarányos terhelése a három változatban

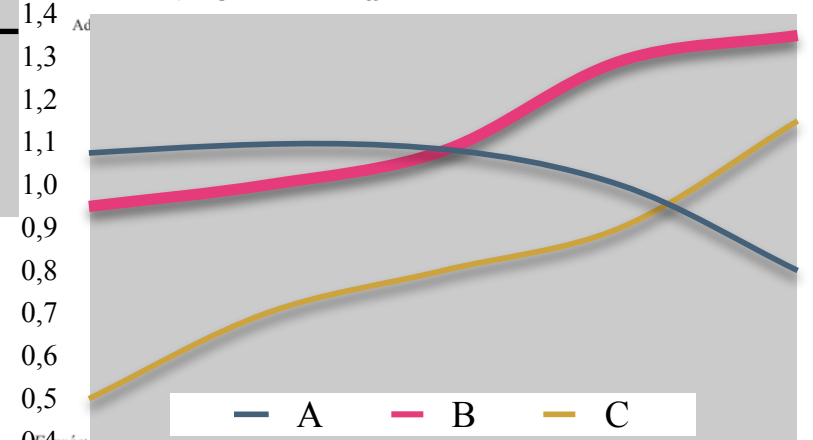
a) a törvény alapján készült értékelési mód szerint



Megjegyzés: lásd a 2. ábra megjegyzését.

Forrás: saját számítás a 2005. évi háztartási költségvetési felvétel alapján.

b) a regressziós modell alapján készült értékelési mód szerint



Forrás: saját számítás a háztartási költségvetési felvétel alapján.

nagyobb terhelés éri, mint az aktív háztartásokat, ugyanakkor ez a különbség a leginkább progresszív esetben összességében eltűnik. A gyermekszám¹⁶ növekedése összességében csökkenti a jövedelemarányos terhelést az A) változatban, de nincs nagy jelentősége a C) változat szerint.

¹⁶ A háztartásban lévő 20 év alatti eltartott gyerekek száma.

	Discovery	Analysis	Writing	Publication	Outreach	Assessment
Trends	social discovery tools	datadriven & crowdsourced science	collaborative online writing	Open Access & data publication	scholarly social media	article level (alt)metrics
Expectations	growing importance of data discovery	more online analysis tools	more integration with publication & assessment tools	more use of "publish first, judge later"	use of altmetrics for monitoring outreach	more open and post-publication peer review
Uncertainties	support for full-text search and text mining	willingness to share in analysis phase	acceptance of collaborative online writing	effect of journal/publisher status	requirements of funders & institutions	who pays for costly qualitative assessment?
Opportunities	discovery based on aggregated OA full text	open labnotes	semantic tagging while writing/citing	reader-side paper formatting	using repositories for institutional visibility	using author-, publication- and affiliation-IDs
Challenges	real semantic search (concepts & relations)	reproducibility	safety/privacy of online writing	globalization of publishing/access standards	making outreach a two-way discussion	quality of measuring tools
Most important long-term development	multidisciplinary + citation-enhanced databases	collaboration + data-driven	online writing platforms	Open Access	more & better connected researcher profiles	importance of societal relevance + non-publication contributions
Potentially most disruptive development	semantic/concept search + contextual/social recommendations	open science	collaborative writing + integration with publishing	circumventing traditional publishers	public access to research findings, also for agenda setting	moving away from simple quantitative indicators

Jeroen Bosman, Bianca Kramer (Utrecht University Library)



■ Követelmények:

- tudományos publikációk !
- alap (raw)?, feldolgozott adatok?
- az adatok strukturális és szemantikai reprezentációi? (metaadat rekordok, kereszthivatkozások, számítógépes modellek adat keresésre, kombinálásra, bányászatra stb.)
- tudományos workflow-k

■ Probléma:

- szemantika feltárása, létrehozása, megőrzése
- szemantikus taggelés

- Probléma:
 - a kutatás minden fázisában keletkeznek RO-ek (teljes kutatási életciklus)
- Követelmény:
 - képesnek kell lenni arra, hogy bármely fázisban tudjunk gyűjteni, eltárolni, kezelni RO-t
- Megoldás:
 - A RO-k aktív életpálya nyomkövetése (digital curation)?

■ Probléma:

- a publikálás aktusa mint egy természetes demarkáció minőség tekintetében is
- RO gyűjtése és eltárolása különféle fázisokból ezt a természetes demarkációt elmossa
- különféle minőségű RO-ek keletkeznek

■ Megoldás:

- RO minőségellenőrzés és minőségbiztosítás szükséges

■ Probléma:

- mely fázisból származzanak az RO-k?
- adatkeletkezés workflow:
 - mért (raw) adat - validálás - szelektálás - kalibrálás - adattranzformáció - adat absztrakció (ad-hoc/standard, gépi/humán eljárásokkal)

■ Probléma:

- RO autentikálása (mint a megosztás, publikálás feltétele)

■ Elvárás:

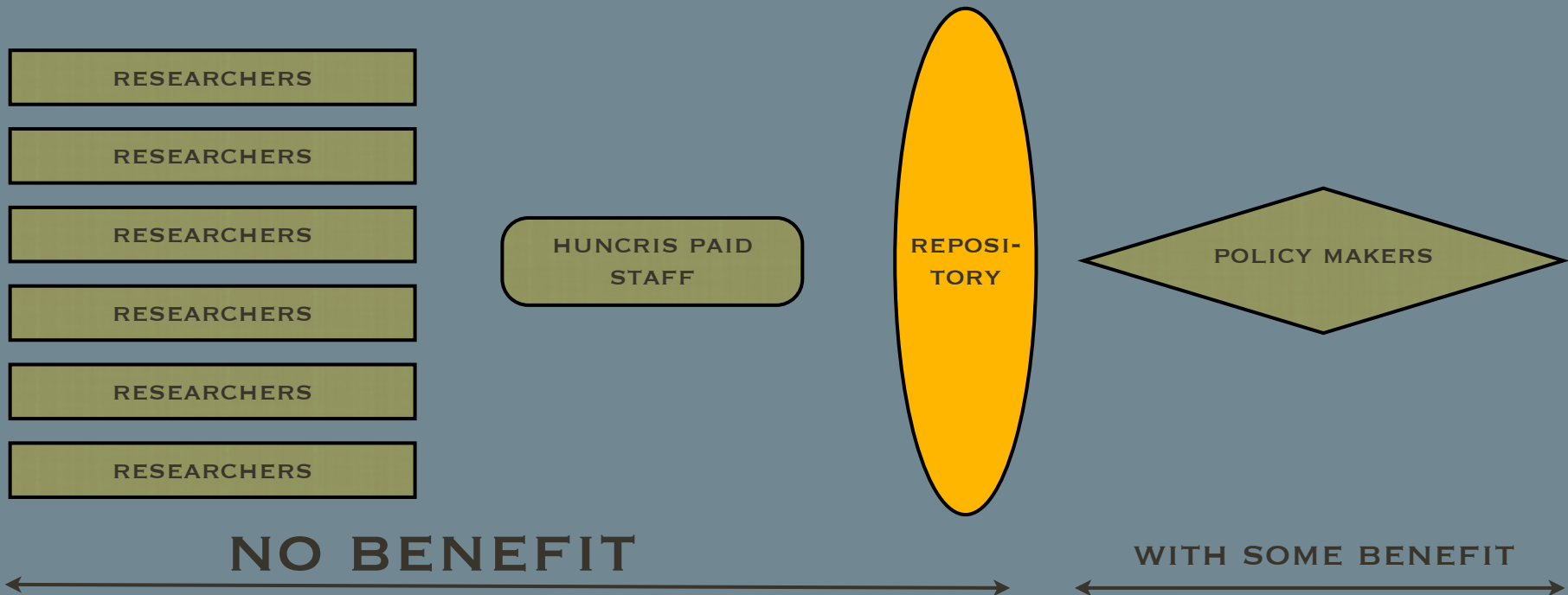
- end-to-end adatkezelés (adatkeletkeztetés, integrálás, megbízható eredet (provenance), hosszú távú (digitális) megőrzés)

- Probléma: a RO értékének meghatározása
 - alapkomponensek értéke, reprodukálás és tárolás költségei
 - nehéz kalkulálni, erősen közösségfüggő
 - sérti a tudomány mertoni normáit
- Probléma: RO kommerciális felhasználása (iparban vagy zárt ipari kutatások esetében)
- Megoldás:
 - AUP (Acceptable Use Policy) - közösségi megoldás
 - a felhasználás (hozzáférés és felhasználás) jogi szabályozása

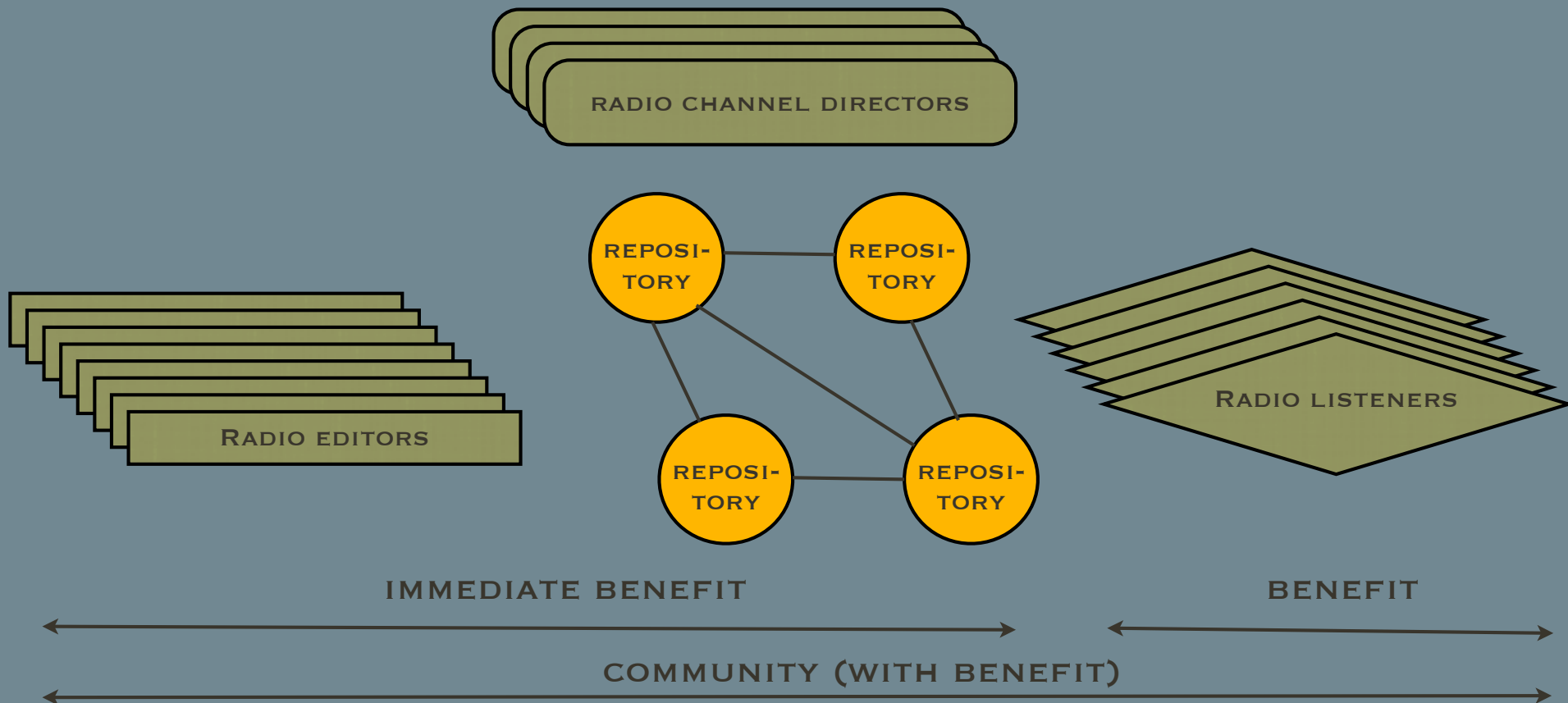
- Probléma: RO felhasználása ipari kutatások esetében (tudományos felhasználás)
- Megoldás:
 - RO piac bevezetése?, (cross border market)?
 - kutatási infrastruktúra támogassa az adás-vétel aktusát?

- Probléma: a RO repozitórium OASYS szabványnak való megfelelése (compliance)
 - Interoperabilitás:
 - repozitóriumok
 - hozzáadott értéket előállító szolgáltatók (service provider) között
- Probléma: Metaadatsémák és applikáció profilok szabványosítása
 - tudományos közösségeken belül
 - tudományos közösségek között
- Probléma:
 - automatikus tudásbegyűjtés (strukturált és nem strukturált tudás esetében)

- Probléma: “Who does the work and who gets the benefit?”
 - Példa: HUNCRIS regiszter (Nemzeti Kutatásnyilvántartás, BME OMIKK)



- Probléma: “Who does the work and who gets the benefit?”
 - Példa: StreamOnTheFly repozitórim hálózat



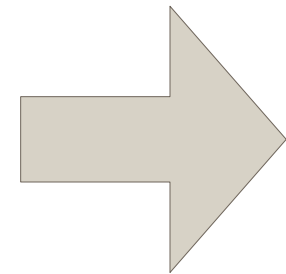
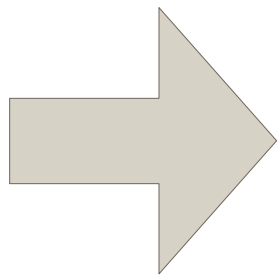
- Probléma: a repozitórium infrastruktúra folyamatos karbantartása és annak költségei
- Elvárás:
 - a előnyök és a munka egyenletes elosztása egyéni, korporációs és közösségi szinteken
 - folyamatos egyéni és csoporttevékenység önkéntes és nem önkéntes alapon

- Probléma: az ipari kutatási eredményekhez való közösségi és publikus hozzáférés kérdése (policy, jogi szabályozás, cégstratégia)
- Probléma: ipari kutatási eredmények hosszútávú megőrzése (közösségi érdek versus cég misszió, költségek vállalása)

- Probléma:
 - mennyiségi
 - minőségi
 - szelekciós mechanizmusok hiánya
- Megoldás:
 - Automatikus információ feldolgozó és tudáskinyerő (szoftver)rendszerek
 - TextMining Center javaslat (MTA SZTAKI)

TextMining Center projekt

Magyar tudomány felszerszámozása a tudományban eleddig még nem járatos eszközökkel (hírszerzés, tudás tömörítés, versenytárs figyelés, újdonság és trendkutatás stb.)



adatállományok
publikációk
hírek, információk

MTMT
tudománymetria
egyéni kutatói értékelés

Magyar tudomány felszerszámozása a tudományban eleddig még nem járatos eszközökkel (hírszerzés, tudás tömörítés, versenytárs figyelés, újdonság- és trendkutatás stb.)




**Magyar tudomány felszerszámozása a tudományban
eleddig még nem járatos eszközökkel** (hírszerzés, tudás
tömörítés, versenytárs figyelés, újdonság- és trendkutatás stb.)



adatállományok
publikációk
hírek, információk

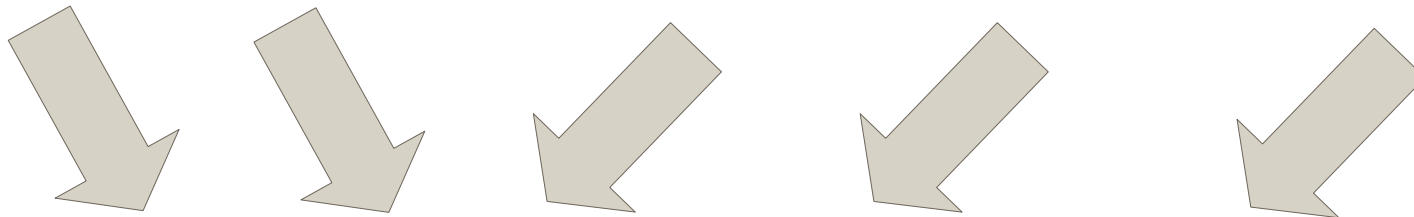
MTMT
tudománymetria
egyéni kutatói értékelés

■ Kutató központ létrehozása text mining célra (SZTAKI)

- szemantikus szintű adat- és szövegfeldolgozás
- a szükséges (intézeti) kutatási infrastruktúra megteremtése (hardver, szoftver, adat infrastruktúra)
- *big data*  *metaadatok, annotációk, szemantikus relációk, rich media stb.*

■ e-Science szolgáltató központ létrehozása (országos)

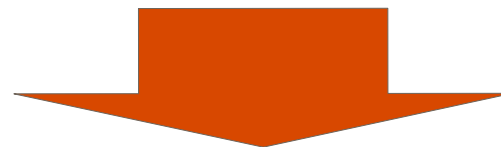
- teljes magyar tudományos közösség számára
- nyílt hozzáférésű internetes szolgáltatás rendszer
- tudásmunkások, kutatók mindennapi (felfedező) kutatási és fejlesztési tevékenységének támogatása
- *új tudás megszerzése, kivonatolása, annotálása, interpretálása, rejtett szemantikus összefüggések feltárása, vizualizálása, megosztása stb.*



**adat fúzió
analízis**

Text Mining Centre

**vizualizáció
megosztás**



open access hálózati szolgáltatások

Kovács László

MTA SZTAKI
Elosztott Rendszerek Osztály

laszlo.kovacs@sztaki.mta.hu

DSD

CONTENTS

-  Projects
-  Services
-  Products
-  References
-  Publications
-  People
-  Contact

MTA SZTAKI

en hu

DSD
DEPARTMENT OF
DISTRIBUTED
SYSTEMS

Copyright © 1994-2013 MTA SZTAKI - DSD.