Annales Univ. Sci. Budapest., Sect. Comp. 43 (2014) 69–87

THE PRAGMATIC NOTION OF INFORMATION

Bernhard Thalheim (Kiel, Germany)

Dedicated to András Benczúr

Communicated by Attila Kiss

(Received June 1, 2014; accepted July 1, 2014)

Abstract. The notion of information is used in a wide variety. It seems that the differences of the notions cannot be integrated. It might also be the case that 'information' is indefinable. The different notions are however used within powerful and rich theories. The theories evolve and thus also the concepts and notions. The notion of information is often implicitly defined in such theories. It is nowadays widely used for terms such as 'information highway', 'information society' and 'digital science'. These terms are oriented towards the human deployment. We thus develop an approach to the notion of information that takes the anthropomorphic characteristics of information in its centre.

1. Revisiting the notion of information

The digital age uses 'information' from a novel point of view [2]. The so-called information highway is mainly a data highway. Whether data become information depends on the point of view the user or the user community has in mind. Therefore we need a novel user-oriented notion. This notion must also consider user communities. Especially due to the development of digital

Key words and phrases: information notion, pragmatics, digital science, communication, knowledge, content, concept, topic, information value, user profile, user portfolio, user life cases

2010 Mathematics Subject Classification: 68Pxx, 94Axx, 03Axx, 03-XX

https://doi.org/10.71352/ac.43.069

science the notion should be robust against the extension to user communities. Information is however also based in this case on data. Data can be transmitted in various ways. They can be discrete or continuous. They are stored in various forms. They influence reasoning and decision making. They can be combined with other data or extracted from data. We use different representations and visualisations. Data have their quality, e.g. they can be more or less precise, more or less biased, more or less reliable etc. All these data may become information for a human. Therefore, we need an approach to the notion of information within this context. We start first with revising the classical notions and develop another notion that could be used in the context of the digital age.

1.1. The syntactic notion of information

Information is certainly one of the most important and fundamental notions. At the same time, we observe a variety of definitions of information. We agree with C. Shannon that it "is hardly to be expected that a single concept of information would satisfactorily account for the numerous possible applications of this general field" [21]. In his understanding information designates surprise, uncertainty, and entropy. Entropy is the measure of uncertainty, of how much choice is involved in the selection of the events or of how uncertain we are of the outcome [20]. This understanding of the notion of information is often used, e.g. [2, 3, 4, 12, 13, 14]. The extreme point of view uses information as what circulate between an emitter and a receiver. We call this notion the *syntactic notion of information*.

1.2. The semantic notion of information

The *semantic notion of information* defines information as a stimuli that has a meaning in some context for its receiver. The Collins dictionary [6] relates information to knowledge*. Knowledge is however a different notion. According to [11], *knowledge is sustainable, evolving, potentially durable and verifiable grounded consensus*. It is in consensus within a world and a community, is based on postulates or principles that create the fundament for the knowl-

*Information is either (1) knowledge acquired through experience or study; (2) knowledge of specific and timely events or situations, news; (3) the act of informing or the condition of being informed; (4a) an office, agency, etc, providing information; (4b) (/as modifier/): information service; (5) (Law); (5a) a charge or complaint made before justices of the peace, usually on oath, to institute summary criminal proceedings; (5b) a complaint filed on behalf of the Crown, usually by the attorney general; (6) (Computer Science) /computing/; (6a) the meaning given to data by the way in which it is interpreted; (6b) another word for data; or (7) too much information.

edge, is true according to a certain notion of 'truth', it is potentially evolving within an ordered evolution/aging process, is reusable in a rule system for new information, it has a longer lifespan and exists with persistent validness, has an effect and is sustaining within a society, community or world, and is not equivalent to other information that can be generated with the aid of facts or preliminary information in the particular inventory of knowledge by a rule system.

Another semantic approach to the notion of information uses operational or combinatorial semantics, i.e. relates the notion of information to some computing device. The amount of information is defined as a function of the number of elements of a finite set within the combinatorial manifold. The algorithmic approach defines the measure of information of an object A in relation to an object B by the length of the minimal program that transforms A from B.

1.3. The pragmatism notion of information

Another notion of information is the *pragmatism notion*: M. Burgin [5, 12] states that information is a "phenomenon that exists in nature, society, mentality of people, virtual reality and in artificial world of machines and mechanisms created by people; information for a system is a capacity to cause changes in the system". Following this approach we can distinguish control mechanisms over information: *personal information public information* and *private information* are regulated by ownership. Information is thus *useful* within a context.

This notion relates information to its deployment. If information is stored in an electronic form it is data. Data is raw material - it has to be processed before it can be turned into something useful. Therefore, information is data that has been processed in such a way as to be meaningful and useful to the person who receives it.

Information is then something that is an objective and abstract entity at the same time. It exists outside and relates to a system. It has, however, no separate existence. Information serves as a basis for adequate behaviour of a system.

This notion is used in business informatics that distinguishes five main uses: for planning, i.e. a business needs to know what resources it has; for recording, i.e. information about each activity or event of interest; for controlling, i.e. for monitoring the plan against the progress and for controlling resources; for measuring success of business; for decision making, i.e. in the form of strategic information for planning the objectives of a business as a whole or in the form of tactical information for decision how resources should be employed or in the form of operational information for check and control whether things are properly evolving.

Contemporary economic research considers information from one side as a resource that provides input into the production process of other commodities and into the control of the market itself and from the other side as an output that is materialised and sold as a commodity.

1.4. Problems of these notions

The syntactic notion of information is content neutral. It is concerned mainly with technical or communicative problems of transmission. Human activity and reasoning uses information for regulation and control. It is not reduction of uncertainty that matters. Information distribution among humans might increase the importance of the information while uncertainty decreases.

The semantic notion of information considers pieces of knowledge. It is based on a known mechanism or calculus for reasoning or one some computing device. In some cases such mechanisms are known and can be effective and effectively used. In most cases however not. The information content for a human is partly determined by the context. Therefore, the reasoning system is going to be based on an open-world assumption.

Information is always somehow encoded. The coding can be known or unknown to a system. The perceptual systems change what information they represent because of the context.

Which role plays information for a human? What is an information user? Should information be measured? Is information consumed? Does somebody give information to somebody else without reducing the own information? How can it happen that somebody gives information to somebody else without having this information?

The notion of knowledge used above suggests a characterisation of knowledge through (1) its content, (2) its concepts, (3) its annotations or topics, and (4) its understanding by the user. Knowledge pieces cannot be considered in an isolated form. For this reason we imagine to use knowledge chunks as a suite of knowledge pieces consisting of content, concepts, topics and information. These dimensions are interdependent from each other. Figure 1 displays the knowledge space.

1.5. The storyline of the paper

A.A. Benczur [2] considered the notion of information for the digital age. We contribute to this work and add the anthropomorphic point of view. When a piece of data becomes information for a human? When it does not? Does the

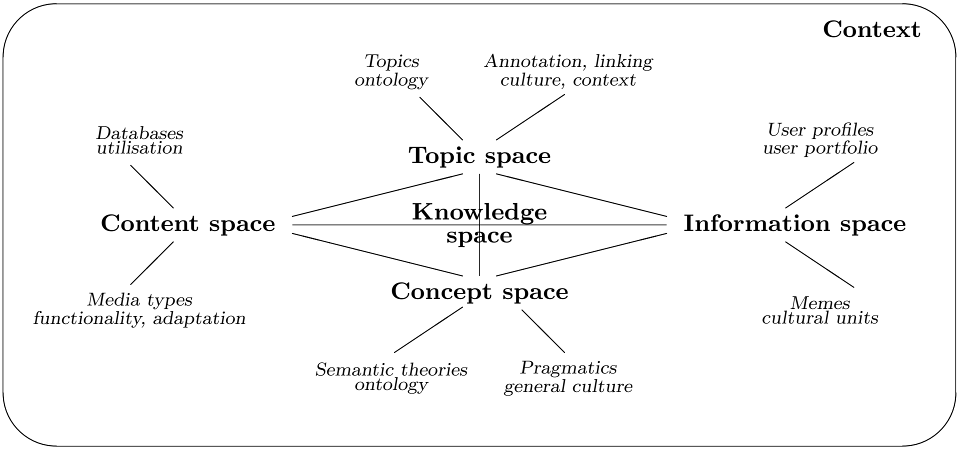

Figure 1. The four dimensions of the knowledge space surrounded by the context dimension: (1) data dimension through content; (2) foundation dimension through concepts; (3) language dimension through topics; (4) user dimension through information; (5) context of data (content) or theories (concept) or user (information) or carrier/language (topic)

ability of processing data have an impact on information receival? Information has a syntactical dimension that can be given through content chunks. It also has a semantical dimension since we want to inform the receiver about something. It also has a pragmatical dimension since we need to annotate, to structure, to illustrate the information. Pragmatics also assumes a meaning of terms by the receiver. Users do not mainly base their utterances on glossaries, thesauri, or ortho-normalised languages. Instead they assume that they will be understood on the basis of context, especially cultural context, their habits, their association to communities or their task background.

The information dimension in Figure 1 is going to be the entry point for a deeper discussion of information in the relationship to a human. We realise that for a human information must be understandable, receivable, interesting and combinable. This observation results in another notion of information.

This notion of information can be used to explain the phenomena in the digital age [2]. We apply the notion to digital science. This novel branch of data-intensive research is based on information services.

2. Pragmatics of information

We may combine the syntactic, semantic and pragmatism notion of information by generalising the GDI notion (General Definition of Information) [10]: *Information is raw data and well-formed and meaningful data that has been verified to be accurate and timely relative to its context, is specific and organised for a purpose, is presented within a context that gives it meaning and relevance, and which leads to increase in understanding and decrease in uncertainty.* "Well-formed" means that the raw data are clustered together correctly, according to the rules (syntax) that govern the chosen system, code or language being analysed. Syntax is understood broadly, as what determines the form, construction, composition or structuring of something. "Meaningful" means that the data must comply with the meanings (semantics) of the chosen system, code or language in question.

This notion does however not entirely reflect the user dimension. Therefore we develop now an anthropomorphic notion.

2.1. The pragmatic notion of information

Definition 2.1. *Information as processed by* humans, *is carried by* data *that is perceived or noticed, selected and organized by its receiver, because of his subjective human interests, originating from his instincts, feelings, experience, intuition, common sense, values, beliefs, personal knowledge, or wisdom, simultaneously processed by his cognitive and mental processes, and seamlessly integrated in his recallable knowledge.*

The *value* of information lies solely in its ability to affect a behaviour, decision, or outcome. A piece of information is considered valueless if, after receiving it, things remain unchanged. For the technical meaning of information we consider the notion used in information theory.

Therefore, information is directed towards pragmatics, whereas content may be considered to highlight the syntactical dimension. If content is enhanced by concepts and topics, then users are able to capture the meaning and the utilisation of the data they receive. In order to ease perception we use *metaphors*. Metaphors may be separated into those that support perception of information and into those that support usage or functionality.

Users are reflected by actors that are abstractions of groups of users. Pragmatics and syntactics share data and functions. The functionality is provided through functions and their representations. The web utilisation space depends on the technical environment of the user. It is specified through the layout and the playout. Layout places content on the basis of a data representation and in dependence of the technical environment. Playout is based on functionality and function representations, and depends on the technical environment.

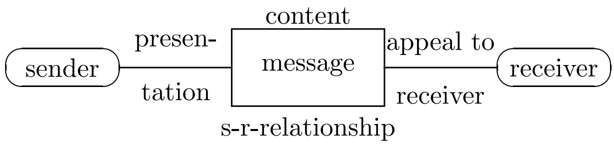

Figure 2. Dimensions of understanding messages

The *information transfer* from a user A to a user B depends on the users A and B, their abilities to send and to receive the data, to observe the data, and to interpret the data. Let us formalise this process. Let s_X denote the function user by a user X for data extraction, transformation, and sending of data. Let r_X denote the corresponding function for data receival and transformation, and let o_X denote the filtering or observation function. The data currently considered by X is denoted by D_X. Finally, data filtered or observed must be interpreted by the user X and integrated into the knowledge K_X a user X has. Let us denote by i_X the binary function from data and knowledge to knowledge. By default, we extend the function i_X by the time t_{i_X} of the execution of the function.

Thus, the data transfer and information reception (or briefly information transfer) is formally expressed it by

$$I_B \;=\; i_B(o_B(r_B(s_A(D_A))), K_B, t_{i_X})\ .$$

In addition, time of sending, receiving, observing, and interpreting can be taken into consideration. In this case we extend the above functions by a time argument. The function s_X is executed at moment t_{s_X}, r_X at t_{r_X}, and o_X at t_{o_X}. We assume $t_{s_A} \leq t_{r_B} \leq t_{o_B} \leq t_{i_B}$ for the time of sending data from A to B. The time of a computation f or data consideration D is denoted by t_f or t_D, respectively. In this extended case the information transfer is formally expressed it by

$$I_B \;=\; i_B(o_B(r_B(s_A(D_A, t_{s_A}), t_{r_B}), t_{o_B}), K_B, t_{i_B})\ .$$

The notion of information extends the dimensions of understanding of message displayed in Figure 2 to a web communication act that considers senders, receivers, their knowledge and experience. Figure 3 displays the multi-layering of communication, the influence of explicit knowledge and experience on the interpretation.

2.2. The communication act

The communication act is specified by

- the communication message with the content or content chunk, the characterisation of the relationship between sender and receiver, the data that are transferred and may lead to information or misinformation, and the presentation,
- the sender, the explicit knowledge the sender may use, and the experience the sender has, and
- the receiver, the explicit knowledge the receiver may use, and the experience the receiver has.

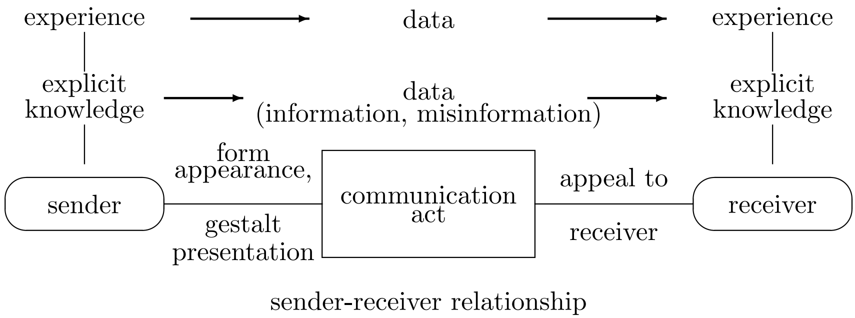

Figure 3. Dimensions of the communication act

We approach the analysis of information system usage as the first important part of storyboarding pragmatics. Information system usage analysis consists of three parts:

1. *Life cases* capture observations of user behaviour in reality. They can be used in a pragmatic way to specify the story space. The work on life cases was reported in a previous publication [19].
2. *User models* complement life cases by specifying user and actor profiles, and actor portfolios. The actor portfolios are used to get a better understanding of the tasks associated with the knowledge system. The work on user models was reported in a previous publication [17].
3. *Contexts* complement life cases and user models by characterising the situation in which a user finds him/herself at a certain time in a particular location. We classify various aspects of contexts related to actors,

storyboard, system and time, which make up the context space, then analyse each of these aspects in detail. This is formally support by lifting relations.

2.3. Profiles and portfolio of the user

User modelling is based on the specification of *user profiles* that address the characterisation of the users, and the specification of *user portfolios* that describe the users' tasks and their involvement and collaboration on the basis of the mission of the knowledge system [16].

To characterise the users of a knowledge system we distinguish between *education*, *work* and *personality* profiles. The education profile contains properties users can obtain by education or training. Capabilities and application knowledge as a result of educational activities are also suitable for this profile. Properties will be assigned to the work profile, if they can be associated with task solving knowledge and skills in the application area, i.e. task expertise and experience as well as system experience. Another part of a work profile is the interaction profile of a user, which is determined by his frequency, intensity and style of utilisation of the knowledge system. The personality profile characterises the general properties and preferences of a user. General properties are the status in the enterprise, community, etc., and the psychological and sensory properties like hearing, motorial control, information processing and anxiety.

A *portfolio* is determined by responsibilities and is based on a number of targets. Therefore, the actor portfolio (referring to *actors* as groups of users with similar behaviour) within an application is based on a set of tasks assigned to or intended by an actor and for which s/he has the authority and control, and a description of involvement within the task solution [18]. A *task* as a piece of work is characterised by a problem statement, initial and target states, collaboration and presupposed profiles, auxiliary conditions and means for task completion. Tasks may consists of subtasks. Moreover, the task execution model defines what, when, how, by whom and with which data a task can be accomplished. The result of executing a task should present the final state as well as the satisfaction of target conditions.

2.4. Users life cases

For task completion users need the right kind of data, at the right time, in the right granularity and format, unabridged and within the frame agreed upon in advance. Moreover, users are bound by their ability to verbalise and digest data, and their habits, practices, and cultural environment. To avoid intellectual overburdening of users we observe real applications before the system development leading to *life cases* [19]. Life cases help closing the pragmatic gap between intentions and storyboarding. They are used to specify the concrete life situation of the user and characterise thus a bundle of tasks the user should solve. Syntax and semantics of life cases have already been well explored in [16].

In addition, each user has an *information portfolio*, which specifies the information needs as well as the information entered into the system. We do not model the information portfolio as part of a user, but instead of this we will model the information "consumed" and "produced" with each more detailed specification of a user request.

3. Towards digital science

The pragmatic notion of information can be extended to user communities. A community obtains a chunk of data and consider this data as information or knowledge if these data can be understood, communicated, processed, and integrated into the body of information or knowledge the community has. This understanding lies behind the development of digital science.

[23] defines the focus and the purpose of digital science as follows: "Digital science focuses on creating an intuitively usable cyber-infrastructure with tremendous capabilities for supporting collaboration and computation. Easy-to-use, human-centered interfaces to cyber-infrastructure created by the digital science will enable the many thousands of researchers in the public and private sectors to use the capabilities of cyber-infrastructure and accelerate innovation and discovery".

Modern science and especially digital science is backed by networks of computing resources, by tools for knowledge management, by ready-to-use and adaptable knowledge. People are at the basis of the future scientific environment. Digital science combines concepts and technologies to make this possible. First emerged in the natural sciences, digital science can be transferred to all disciplines and benefit from their integration, including the arts and humanities. The adoption of digital science has only just begun; technologies, organisational structures and culture progress alternately towards the digital science vision.

[7] defines digital science through their backgrounding technology, especially within an yet-another-edutainment (or -e-learning) platform: "Digital science

will focus on providing world-class software tools and services to scientists, managers and funders with the ultimate aim of making research more productive through the use of technology." This technology supports the delivery

- of *information* to my community within the current context,
- of *lectures* together with settled material, and
- of *competencies* gained by each of the partners.

Digital science (sometimes called e-science or science 2.0) is characterised by

1. collaboration (or crowd) stories for development of science within a community (how),
2. profiles, level of engagement and interest of collaborators (who),
3. content that is either private or shared based a number of sharing pattern (what),
4. coherent presentation of content depending on the profile of the user and on progress of work within a community,
5. rights, roles and plays of contributors within the story assigned (which rights portfolio), and
6. constraints for the participation and contribution (which conditions).

3.1. Information services for digital science

Digital science relies on high availability of content at any time for convenience (when), at any place of the collaborator (where in the 'cloud'), and within any context that is agreed in advance. Digital science is based on the classical science behaviour, classical science standards, classical editing of scientific results and classical quality norms and assurance (communalism, universalism, disinterestedness, originality, and skepticism) within a scientific community.

Digital science is science that extensively uses digital (or IT) services. These services might be

- *services for collaboration* within a group or community of scientists,
- *data, information or knowledge services* for collection, creating, delivering, maintenance and cleansing of content within a scientific community,

- *computational, exchange and control services* for the portfolio of tasks a scientist has been assigned,
- *network services* for hooking into the network at any time independently of the current location and environment of the scientist, and
- *protection services* for privacy and security of scientific communities.

Services for digital science require a more sophisticated description for services beyond WSDL. We use the following frame for the specification of digital science services and for models in general [22]:

End (wherefore) of the services;

Sources (whereof) used for the service;

Supporting means (wherewith) the service is relying on;

Surplus value (worthiness) the service might give to the users;

Purpose (why, whereto, when, for-which-reason) of the service;

Activities (how) supported by the service for collaboration stories based on data consumption (what-in) and resulting in data production (what-out);

Parties such as suppliers (by-whom), consumers (to-whom), and producers (whichever);

Application domain describing the application area (wherein), application cases (wherefrom), the problems (for-what), the organizational unit (where), triggering events (whence), and IT data, control, computation (what, how)

Context for the service such as the system context (whereat), the story context (where-about), the coexistence context (whither), and the time context (when).

Digital science takes its cue from the technologies of Web 2.0, Web 3.0 and Knowledge Web. It creates conversations between researchers, lets them discuss the data and connect it with other data that might be relevant.

The most important property of digital science is its constant collaboration among scientists. Collaboration of partners consists of communication, coordination, and cooperation. Cooperation is based on cooperativity, i.e. the disposition to act in a way that is best helpful for the collaboration partners, taking their intentions, tasks, interests and abilities into account. At the same time, collaboration is established in order to achieve a common goal. Actors choose their actions and organise them such that their chances of success are optimised with respect to the portfolio they are engaged in. Additionally, the social context may be taken into account, which consists of interactive and reactive pressures. Typical social enhancements are socially indicated situations such as welcome greetings, thanking, apologising, and farewell greetings.

Digital science services support a scientific community. This community is backed by a common library of this community. The essential results of such communities are knowledge chunks compiled. Communities collaborate. Therefore, exchange services are an essential element of the knowledge infrastructure. The collaboration within the community must be supported corresponding services. We thus may separate services that support the infrastructure of digital science as given in Figure 4 [11].

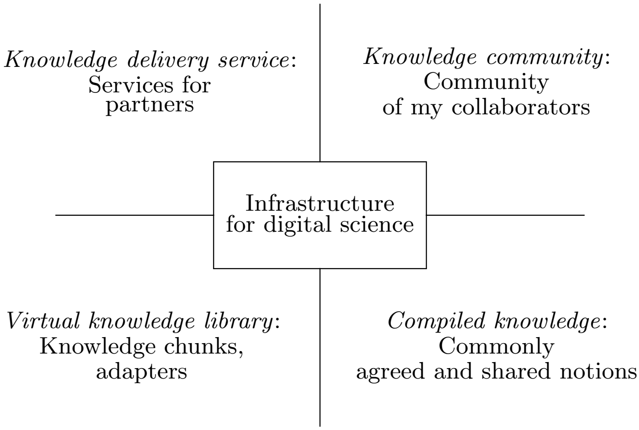

Figure 4. The dimensions of infrastructures for digital science

3.2. Information infrastructure services for digital science

Information infrastructure services support collaborative work of scientists independent from their location at different periods of time. The collaboration is based on roles, rights and group membership of scientists. These services must support the entire research process - starting from brainstorming, detection, discussion continuing through issuing project and should not terminate after a project has been closed. Data that are bundled with research results must be accessible together with publication of results.

Virtual research environment services thus provide an access to real science resources (e.g., data analysis and computation tools, data). They thus support the scientific work of generation Z. Young people are nowadays used to the facilities of the internet, to the richness of data sources, to the variety of beliefs and opinions displayed at pleads of webpages, to the constant change in the web and to the incompleteness and misinformation of opinion pools and blogs. The generation Z has however not yet found a way to discover the real information, to dig to grounded data, to follow links to the consensus that has led to the data in the web, and select correct and consistent data within the world of data in the web. Therefore, a number of tasks must be solved for digital science:

Digital archives beyond the classical hosting and archive technology must be backed by stabile and dependable access services, simple but powerful search facilities, and high-performance data delivery. They ensure the long-term availability of digital media and contents that have been acquired from around the world and their integration in the digital research environment. Digital archives must be backed by subject-specific standards and methods of data curation and archiving.
Archives are either [1]

- *light archives* that can be accessed by many authorised users,
- *dark archives* that cannot be accessed by any current users but may be accessible at future dates subject to the occurrence of specific pre-defined events either of type 1, i.e., providing a form of escrow or "bit preservation" of content, or of type 2, i.e., providing the bit preservation of the content plus some degree of associated services for future access, and
- *dim archives* that provide bit preservation for the content plus digital preservation planning and actions for long-term perpetual access, and also limited current access.

Library 2.0 tools guarantee the broadest possible access to digital publications, primary research data, to virtual research and communications environments, and other material without unexpected costs and other barriers. They ensure the systematic backup, archiving and provisioning of scientific data for subsequent (re-)use by third parties. Perpetual access guarantees to the right of the subscriber and their users to have ongoing permanent access to electronic materials which have already been leased and paid for by the subscriber from a publisher.

Search and access tools are currently based on database techniques for ad-hoc query formulation and computation. Instead knowledge infrastructures must be backed by support for access in dependence on meta-data about the users, their profile and portfolio, and especially their information demand. The same set of requirements is valid for search tools. Search can be categorised into seven categories [8]. This categorisation allows a development of standardised methods for access to information massives.

Collaboration tools for collaborative research within continuously changing communities would allow broad cooperation of scientists, would support concentration of competencies and resources and would improve coordination of current and future activities. Collaboration services can be build based on the 3C framework [18] that separates supporting services into communication services, coordination services, and cooperation services. These services can be supported by current database technology [9].
Since science data are heterogeneous in their nature a common data policy must be promoted. This policy is based on a *data management plan* for all collaborating partners. This plan includes maintenance procedures, clearing procedures, control of access and data consumption, data update procedures for all partners, and data delete procedures for all partners. Such policies must be the basis for data collaboration contracts. At the same time, science data must be bundled with the publication of results drawn from them.

Distribution tools provide ideal conditions for delivery, distribution, and reception of most recent research results.

Integration tools support the constant change of the world of knowledge which is continuously extended, modified and documented in scientific publications.

3.3. SWOT analysis of the current state-of-the-art

Knowledge management is a long-standing research issue, was probably the hottest topic in management in the 1990s and has partially failed. We may learn the lessons for the development of infrastructures that support digital science. Despite massive investments and a lot of highly motivated people, the best knowledge management systems succeeded at capturing and institutionalising the knowledge of a company. The real value of knowledge management is in creating new knowledge, rather than simply “managing” existing knowledge.

What we really need are new approaches to creating knowledge, ones that take advantage of the existing software systems and incorporate new digital infrastructures for intensive collaboration and for mobilisation of benefits of interacting and collaborating communities. Digital science is going to heavily rely on shared network platforms, provides tools and forums for knowledge creation while at the same time capturing the discussion, analysis, and actions in ways that make it easier to share across a broader range of participants. Blogs and wikis have great benefits for a community and any member of such.

At the same time 'knowledge management' was really a misnomer. Most companies meant information management and called it knowledge management. They captured tacit information held by the employees and thus made it explicit. Collaboration was not yet a target. Employees did not see added value to their workflow by providing information to the company without any really effective enhancement to the workflow. They will participate when there are direct benefits. People want knowledge in exchange for their information. Current tools successfully collect information but are very poor at yielding knowledge.

Digital science services focus on providing immediate value to scientist in terms of helping them tackle difficult performance challenges while at the same time reducing the effort required to capture and disseminate the knowledge created. They bring into play network effects in the generation of new knowledge. They leverage the social aspect in worthwhile ways.

Currently, many systems are available for digital science due to the large number of projects such as the AstroGrid, Avian Flu HSN1, Comb-e-Chem, DiscoveryNet, the European DataGrid (EDG), the gSLM, the GriPhyN, Indonesian Earthquake, WISDOM projects and the EGI, GEANT, Large Hadron Collider (LHC) at CERN, NERC DataGrid, myGrid, PRACE, and RealityGrid grids. Based on the experience gained different e-Business, e-Government, and e-Services have been developed. e-Science project can benefit from experience and analyses and can share the services that have been.

We may use SWOT analysis techniques for a characterisation of the current state of the art:

Strengths and goodies: Basic technologies are on hand. They are mainly based on well-known techniques of "programming in the small". They use partially "programming in the large".

Weaknesses and missing IT: collaboration tools for crowdworking Techniques for "programming in the world" are chaotically introduced and must be matured and systematised.

Opportunities of current technology: The world-wide collaboration on the fly allows to use human and technical resources wherever they are, whenever they are developed, whoever can provide the correct data at the best point of time, in the agreed format and quality for the right user with the at the right location and context.

Threats of current technology and restrictions: The chaotic development uses a wide variety of techniques, does not yet have a basement and is thus difficult to evolve. Instead we need flexible frameworks and a thoughtful integration of technical knowledge that has been obtained in the past. Data (sometimes called information) are typically delivered without quality meta-information or facilities for information management. The search is still address-based and looks alike the Computer Stone Age. Programming in the world is a common practice, however mainly in a quick and dirty fashion.

4. Conclusion

The notion of information is certainly one of the most overloaded and diverse in natural sciences and engineering. We distinguish between syntactic, semantic and pragmatism notions of information and realize that these notions do not reflect the relation between the data and the human. Therefore, a pragmatic notion of information is introduced in this paper. This pragmatic notion relates data to the users, their interests, their current focus and scope, their abilities to digest these data and their capacity to use this data, e.g. for integration into their knowledge space. The syntactic notion mainly reflects the concept space. The semantic notion also considers the concept space. The pragmatism notion considers information as a utility. The pragmatic notion however considers the information space of a user.

This notion also enables to handle data within digital science (or e-science, science 2.0). Digital science uses data for communities of users. Therefore, we may extend the information notion to communities instead of considering one singleton user. The commodity of data is ascertainable through the consideration of the data as information by a singleton user and must be supported by tools that enable users to consider the data chunk as an information chunk within this community. Services provide one kind of support for such communities. They must however be combined with some corresponding infrastructure.

The user community changes however due to the development of digital science. Digital natives read code and data better than natural language. This change of usage and deployment can be observed nowadays. The pragmatic notion of information will become the central notion of information within this setting.

References

[1] **Beagrie, C.**, Ensuring perpetual access, `http://www.allianzinitiative.de/fileadmin/hosting_studie_e.pdf`

[2] **Benczúr, A.**, The digital universe - an information theoretical analyses, *CompSysTech*, ACM, 2013, 1-10.

[3] **Benczúr, A.,** Information measurement in relational databases, *MFDBS*, LNCS **305**, Springer, 1987, 1-9.

[4] **Benczúr, A.**, The evolution of human communication and the information revolution - A mathematical perspective, *Mathematical and Computer Modelling*, **38** (7-9) (2003), 691–708.

[5] **Burgin, M.**, *Theory of Information. Fundamentality, Diversity and Unification.* World Scientific Publishing Co, 2010.

[6] **Collins English Dictionary.** Complete and Unabridged, `http://www.thefreedictionary.com/_/misc/-HarperCollinsProducts.aspx?English`

[7] **Digital science community**, `http://www.digital-science.com/`

[8] **Düsterhöft, A. and B. Thalheim**, Linguistic based search facilities in snowflake-like database schemes, *Data and Knowledge Engineering*, **48** (2004), 177–198.

[9] **Fiedler, G., T. Raak, and B. Thalheim**, Database collaboration instead of integration, *APCCM'05*, CRPIT **43**, Australian Computer Society, 2005, 49-58.

[10] **Greco, G.M., and G. Paronitti and M. Turilli and L. Floridi**, The philosophy of information: A methodological point of view, *Wissensmanagmeent*, DFKI, Kaiserslautern, 2005, 563-570.

[11] **Hashimoto, T., J. Henno, H. Jaakkola, A. Sasa and B. Thalheim**, Infrastructures for knowledge systems environments, *Information Modelling and Knowledge Bases XXIII*, Frontiers in Artificial Intelligence and Applications **237**, IOS Press, 2012, 369-398.

[12] **Henno, J.**, Emergence of language: Hidden states and local environments, *Information Modelling and Knowledge Bases XIX,* eds. H. Jaakkola, Y. Kiyoki and T. Tokuda, IOS Press, Amsterdam, 2013, 170-181.

[13] **Kolmogorov, A.N.**, On the Shannon theory of information transmission in the case of continuous signals, *IRE Transactions on Information Theory*, **2** (4) (1956), 102–108.

[14] **Kolmogorov, A.N.**, Three approaches to the quantitative definition of information, *Problems of Information Transmission*, **1** (1) (1965), 1–7.

[15] **Kolmogorov, A.N.,** Logical basis for information theory and probability theory, *IEEE Transactions on Information Theory*, **14** (5) (1968), 662–664.

[16] **Schewe, K.-D. and B. Thalheim**, *Usage-based storyboarding for web information systems,* Technical Report 2006-13, Christian Albrechts University, Institute of Computer Science, Kiel, 2006.

[17] **Schewe, K.-D. and B. Thalheim**, User models: A contribution to pragmatics of web information systems design, *Web Information Systems – Proceedings WISE 2006*, eds. K. Aberer, Z. Peng and E. Rundensteiner, LNCS **4255**, Springer, Berlin, Heidelberg, 2006, 512-523.

[18] **Schewe, K.-D. and B. Thalheim**, Development of collaboration frameworks for web information systems, *20th Int. Joint Conf. on Artifical Intelligence, Section EMC07 (Evolutionary models of collaboration)*, Hyderabad, 2007, 27-32.

[19] **Schewe, K.-D. and B. Thalheim**, Life cases: A kernel element for web information systems engineering, *Web Information Systems and Technologies*, Lecture Notes in Business Information Processing **8**, Springer, Berlin, Heidelberg, 2008.

[20] **Shannon, C. E.**, A mathematical theory of communication, *Bell System Technical Journal*, **27** (1948), 379–423, 623–656.

[21] **Shannon, C. E.**, *Collected Papers*, IEEE Press, 1993.

[22] **Thalheim, B.**, The conceptual model ≡ an adequate and dependable artifact enhanced by concepts, *Information Modelling and Knowledge Bases*, vol. XXV of *Frontiers in Artificial Intelligence and Applications, 260*, 241–254, IOS Press, 2014, 241-254.

[23] **The Digital Science Center at Indiana University**, http://pti.iu.edu/dsc

Bernhard Thalheim
Department of Computer Science, Christian-Albrechts-University Kiel,
24098 Kiel, Germany
thalheim@is.informatik.uni-kiel.de