

CAN YOU BELIEVE WHAT YOU READ? GUARANTEEING TRUSTWORTHY SEARCH RESULTS AND CONTENT DISCOVERY

A brief report of the third session at NGIForum18: “Better search for trustworthy content and objects discovery” by SpeakNGI.eu. Porto 13th September 2018



Figure 1: Shots from the “Better search for trustworthy content and objects discovery” interactive session at NGI Forum 2018



CAN YOU BELIEVE WHAT YOU READ? GUARANTEEING TRUSTWORTHY SEARCH RESULTS AND CONTENT DISCOVERY

A brief report of the third session at NGIForum18: “Better search for trustworthy content and objects discovery” by SpeakNGI.eu. Porto, Portugal, 13th September 2018

Moderators:

Alessandro Bassi
Alexandru Stan
Jamal Shahin
Aleš Černivec

Rapporteurs

Paul Malone
James Clarke
Sara Pittonet
Andrea M. Schillaci

Semantic Data Organisation

Alessandro Bassi of ABC France led this discussion about **Semantic Data Organization** and the animation was carried out by Paul Malone, SpeakNGI.eu project.

Alessandro firstly presented a position on how we need new ways of storing, understanding and releasing data based on the semantics of the data itself and the context within which the data exists.

"We have many options how to store our data (cloud, device, edge, NAS). There is a large amount of data that once created can be distributed. But is it better to rely on external devices or buy a cloud space somewhere? Imagine you have a heart attack and you need people to access your medical history, immediately: where would you like them to go?"

Computer and human memory are working in a totally different way. We don't know where things are archived in our brain.

What if we could access medical data in real time, according to a context? This is technologically possible, but which kind of privacy should apply? Can we have a software licence type scheme also for data?

There was then a discussion around a flip chart designed to gain an understanding of the topic and categorise the human values encountered with the topic linked with the challenges, potential solutions, known initiatives addressing those solutions and also gaps and R&I needs that can successfully provide solutions that address the challenges and plug the gaps. The identified relationships are captured in Figure 2, below. The same process was repeated for all the topics of the session. The relationship between the values, challenges, potential solutions, etc. is shown in square brackets indicating how these concepts for Semantic Data Organisation relate to one another, in a systematic way.

This discussion evolved from the idea that our current means of storing and releasing data lacks any insight into the meaning of that data, the context in which it is in use and reliable set of rules governing how that data can be used or shared given that context in addition to automated mechanisms to enforce those rules.

The following Human Values were identified in relation to this topic ->
V1. Privacy V2. Trustworthiness
The following Challenges were identified in relation to these Human values ->
C1. Contextual awareness of data [V1, V2] C2. Technology challenge: Organising Data through meaning (Files Systems) [V2] C3. Automatic Extraction of Metadata [V2] C4. Cloud vs Location assurance [V2] C5. Device Data vs Personal Data [V1, V2]
The participants identified a number of Potential Solutions to address the challenges ->
S1. AI to understand context [C1, C3] S2. Sticky Policies [C4, C5] S3. Data Licencing [C1, C5] S4. Personal Data Rules [C4]
A number of known initiatives / projects were identified addressing the solutions ->
I1. CocoCloud project [S2] I2. Creative Commons licencing [S3] I3. Picos [S4] I4. Personal Data Vaults [S4]
Identified Gaps ->
G1. AI to extract semantic meaning [P1, C2, C3] G2. Education of Creative Commons plus extensions to CC type licencing [S3] G3. Solutions to contextually manage data release [S1, S4]
R&I Needs to fill the gaps ->
R1. Remote / Automated enforcement of data handling rules [G3, S1]

Figure 2: Human Values, Challenges, Solutions, Initiatives, Gaps and R&I needs related to Semantic Data Organisation

SME perspectives for trustworthy search and content discovery.

The session was moderated by Mr. **Alexandru Stan**, In-Two, Germany and animated by James Clarke, SpeakNGI.eu project.

Background: Mr. Alexandru Stan was contacted through the involvement of In-Two within the [NEM Technology Platform](http://www.nem-initiative.org/) (<http://www.nem-initiative.org/>), which offers a platform to European SMEs working on media and content-related technologies.

The talk provided the viewpoint of an SME working on developing content-rich applications with trustworthiness and quality of content, as well as provides powerful search capabilities and content exploration interfaces. Some of their applications relate to events, tourism and cultural heritage. The presentation focused on a number of key points for SMEs in relation to their participation to the Next Generation Internet initiative:

1. Are **SMEs** important for NGI?
2. Can SMEs do it **alone**?
3. There is **value** in collaboration!

In response to the first question, it was felt that SMEs do not play well alone!

As stated by Mr. Stan, "*Researchers need the collaboration with SMEs in order to make their innovation an industrial success. Industry can push the innovation process behind research organizations*". Mr. Stan highlighted the Marconi project (<http://www.projectmarconi.eu>) for an example of excellent collaboration between research and industry, in particular SMEs, within the combined fields of radio & media. IN2 is an SME working on developing content-rich applications with trustworthiness and quality of content, and collaboration with research institutes; for them, the results are very stimulating, according to Mr. Stan.

The importance of the role of SMEs in the NGI initiative was especially highlighted, since they have a critical role to play in the value chain of innovation, in which they are constantly experimenting, and rapidly bringing new innovative things to the market. They are well positioned and have a guided view on how to design solutions that are "People centric" and "human centric", as to succeed as an SME, it is necessary to come up with real solutions that make sense to solving real world problems.

The following Human Values were identified in relation to this topic ->

- V1. Openness;
- V2. Access;
- V3. Inclusiveness;
- V4. Trustworthiness.

The following Challenges were identified in relation to these Human values ->

- C1. The trend of moving towards echo chambers and walled gardens, and how this differs from the ideas of an open internet [V1, V2, V3, V4];
- C2. Content without the proper context or content with the wrong context [V4];
- C3. The need to raise the profile and reputation of SMEs producing solutions in the larger setting against the large players, who have a dominant role [V3];
- C4. Coming up with a peer review or reputation strategy to make it clear to customers that what SMEs are doing is trustworthy [V3, V4]
- C5. Trustworthiness for the entire value chain in large scale systems – how to ensure all steps along the way are verifiable in real time (some examples from smart Agriculture, smart pharma were highlighted where if one sensor was giving incorrect readings, the whole system is compromised) [V4].

The participants identified a number of Potential Solutions to address the challenges ->

- S1. Less is More (e.g. "A-Social networks");
- S2. Conversation vs. "Engagement";
- S3. Publish (on your) Own Site, Syndicate Elsewhere (IndieWeb);
- S4. Trade off: Convenience / Privacy;
- S5. Selective serendipity (i.e. giving the algorithms an opportunity to step out of the models);
- S6. Developing transparency into peer to peer reputation systems, making them more trustworthy – connect with experimental platforms such as smart cities initiatives, Fed4Fire, and other experimental facilities would enable raising of profile and reputation of results coming from the NGI programme [C4].

A number of known initiatives / projects were identified addressing the solutions ->

<ol style="list-style-type: none"> 1. Caprice community: https://www.caprice-community.net/ 2. European Trusted Cloud Platform of EIT Digital; e.g. "My Data Store" of Telecom Italia (https://www.slideshare.net/CDECatapult/4-my-d-atastoretelecomitalia). Also see Trusted Cloud Platform with FSecure, experimenting with an offer of Security as a Service (now available commercially here: https://aws.amazon.com/marketplace/pp/B01N3NNLIW?qid=1536780833483&sr=0-1&ref=srh_res_product_title). In-Two have integrated this in one of their solutions so that our customers could have secure and trustworthy content hubs, building search and discovery on top of the trusted content only. [C5] 3. Pharma industry are working on some elements of the E2E verification process but needs more R&I [C5, S6]; 4. FIWARE context broker, which is a central component of the FIWARE platform, would it address the challenge related to mapping the content and context in relation to the content. [C2]
<p>Identified Gaps -></p> <ol style="list-style-type: none"> G1. How to make sure the context for the content is correct and evolving? [C2, I4] G2. Lifespan of the content and whether it is evolving according to the correct context? Some challenges to address are a) Gaining reputation from today's models could take too long; b) Need to develop new ways to quicken the process e.g. having reputable review trusted authorities and/ or certification processes. [C2] G3. Need to look at the end to end harvesting of data to ensure there is no corruption of data anywhere. [C5, S6]
<p>R&I Needs to fill the gaps -></p> <ol style="list-style-type: none"> R1. Detailing correctness and authenticity of entire value chain verification mechanisms (some work started in smart pharma and smart agri, but much more is needed for NGI); [G3] R2. Could blockchain technologies be used to address these gaps in such a low power based system: is it feasible, either technically or economically? [G1 – G3]

Figure 3: Human Values, Challenges, Solutions, Initiatives, Gaps and R&I needs related to Search and Discovery (SME perspective)

Tools and Concepts for search and Discovery based on the Slovenian Interoperability Framework

Mr. **Aleš Černivec**, Project Manager, XLAB d.o.o. in Slovenia led this discussion session and the animation was carried out by Mr. Andrea M Schillaci of the SpeakNGI.eu project.

Background: Aleš Černivec is a member of the NGI European Champion Panel (ECP) and he has been leading implementation of Slovenian interoperability portal (<https://nio.gov.si>) and contributed to Slovenian open data portal (<https://podatki.gov.si>). Mr. Černivec presented on search and discovery tools from the perspective of the Slovenian interoperability and open data portal, in particular covering security, ID management, authentication and authorisation frameworks, data sharing and safeguarding privacy.

XLAB d.o.o. in Slovenia, a Slovenian company, has gained significant experiences in building knowledge management services and portals, including a national interoperability portal that allows public authorities (PAs) to publish their solutions and make them discoverable to citizens. Legislation and openness of public sectors data allows citizens to look in to PAs operations and better trust them. Citizens have the means to develop new solutions based on searchable and existing information at their ready disposal.

“Openness - management - trust - security - revision - responsibility are the keywords for these environments”, according to Mr. Černivec!

The development of the open data portal addresses legislation on openness of public sector’s data, information and solutions. The solutions developed have a number of advantages, enabling the citizens a look into the operation, thus increasing trust and revision, makes possible the development of new solutions based on the data, the institutions producing the data become more aware of their responsibility leading to the improvement in data quality, quantity and management, and it provides a single point of access for easier search, analysis and linking of the data.

The presentation focussed on the technologies involved and also highlighted the human values to be considered while developing these technologies. The human values of importance to this topic include openness, trust, security, revision (transparency), and responsibility. The processes developed are demanding knowledge in relation to search, analysis, linking, and (knowledge) management.

Mr. Černivec concluded his talk by stating, *“Experiences gained show that the technology allows us to manage the knowledge easier and faster, produce new content faster; and human values can now be powered by the technology and Next Generation Internet should make knowledge management easier, more dynamic, discoverable.”*

The following **Human Values** were identified in relation to this topic ->

- V1. Openness
- V2. Trust
- V3. Security
- V4. Transparency
- V5. Responsibility
- V6. Inclusiveness
- V7. Awareness

The following **Challenges** were identified in relation to these Human values ->

- C1. Legislation on openness of public sector data [V1, V4]
- C2. Increasing trust and revision [V2]
- C3. People engagement [V7]
- C4. Improvement in data quantity and quality and management [V5]
- C5. Education [V5, V7]
- C6. Advantages on the market (for small and big players) [V6]
- C7. Fragmentation (objects) [V1, V2, V3, V4, V5]

The participants identified a number of **Potential Solutions** to address the challenges ->

- S1. Distributed Ledger Technologies (DLTs) [C1-C4, C7]
- S2. Certifications [C2, C3, C5]
- S3. Excluding blockchain [??]
- S4. Micro-services oriented solutions [C2, C6]

A number of known initiatives / projects were identified addressing the solutions ->
I1. Public Service Initiative (PSI) [S2]
I2. NIFO [V4, C2]
I3. JoinUp https://joinup.ec.europa.eu/
I4. Open Data Portal https://data.europa.eu/euodp/en/home
I5. ISA2
I6. My Data Hub
Identified Gaps ->
G1. Usability, Accessibility [V1, V3]
G2. Business Models on data [C6]
G3. Transparency on retaining data [V4, C1, C2]
G4. Skills (HR) [V1, V5, V7, C3, C5]
G5. Sources of funding [C6]
R&I Needs to fill the gaps ->
R1. DLTs [V1 – V4]
R2. Survey supporting NGI [V1,V4,V6]
R3. Consultation platforms for citizens [C3]
R4. Semantic search [V4]
R5. Specialised Search (IoT, Big Data, AI, ...) [G5]

Figure 4: Human Values, Challenges, Solutions, Initiatives, Gaps and R&I needs related to Interoperability framework and open data portals

Tackling online disinformation: a European Approach

Prof. **Jamal Shahin**, VUB and GIPO, led this discussion session and the animation was carried out by Ms. Sara Pittonet of the SpeakNGI.eu project.

Background: Prof. Shahin is a member of the NGI Champions Panel and also was involved in animating The Global Internet Policy Observatory (GIPO), which has now been passed onto the SpeakNGI.eu project. The continuing work in GIPO is geared towards proactively bringing together both the NGI Policy and Technology communities.

Prof. Shahin presented an overview of challenges following recent ‘disinformation’ crises. Three highlighted challenges include: skills (for the public), public sphere (media/journalism), platforms and ‘algorithms’.

Prof. Shahin asked the audience, *“Did you know that Titanic did no sink; it burnt?”* This was the subject of a Youtube video that was listened to by some children as it was *“on youtube, so it must be true!”* This begs the question: *“What critical skills do we need to have to overcome disinformation?”*; and *“What tools do we need to overcome disinformation in the NGI?”*

Prof. Shahin presented some statistics that were quite illuminating and disquieting:

- Only 2% of children have the critical literacy skills they need to tell if a news story is real or fake;
- Half of teachers (53.5%) believe that the national curriculum does not equip children with the literacy skills they need to identify fake news.

A number of challenges related to tackling online disinformation were presented, including challenges related to existing online platforms – where is the liability / who takes the responsibility; What are the tools and techniques required for overcoming disinformation; Skills - digital literacy; and how to reach out into the Public sphere and the news media ecosystem.

Prof. Shahin involved the audience in an interesting poll about how to face online disinformation. A number of answers were recorded from the participants on the following possible solutions:

- Through self-regulation (5.2 / 10, where 10 is strongly agree);
- Through support to quality journalism (7.7 / 10);
- Through regulation (5.0 / 10);
- Through education (8.7 / 10);
- Non-intervention – the marketplace of ideas will deal with it (2.8 / 10).

The following Human Values were identified in relation to this topic ->
<ul style="list-style-type: none"> V1. Authenticity V2. Authoritative V3. Diversity V4. Transparency V5. Trustworthiness
The following Challenges were identified in relation to these Human values ->
<ul style="list-style-type: none"> C1. We are living in a world of sensationalism and negativity [V1, V2, V4, V3] C2. Advertising does not equal content trustworthiness (e.g. google a product and you may find a review that is a hidden advertisement disguised a review) [V1, V2, V4, V5] C3. Creation of needs is still largely based on or driven by large entities [V3, V5] C4. Authority of the platform [V1, V2, V5] C5. Information overload [V3] C6. Open / Democratised Algorithms [V3, V4]
The participants identified a number of Potential Solutions to address the challenges ->
<ul style="list-style-type: none"> S1. Reputation meter for ranking information [C1, C4, C6] S2. Control (at certain times) [C6]
A number of known initiatives / projects were identified addressing the solutions ->
<ul style="list-style-type: none"> I1. Global Internet Policy Observatory (GIPO) I2. Fact-checking organisations [S1]
Identified Gaps ->
<ul style="list-style-type: none"> G1. Education [C1] G2. Integration between technology and human intervention [S1, S2] G3. Means of control [S2]
R&I Needs to fill the gaps ->
<ul style="list-style-type: none"> R1. Supervisor (AI supported) [I2] R2. Who is addressing the human / social effect

Figure 5: Human Values, Challenges, Solutions, Initiatives, Gaps and R&I needs related to Tackling online disinformation – an EU approach.

Conclusions

The session on **Better Search For Trustworthy Content And Objects Discovery** has identified a number of important research and innovation (R&I) topics for the NGI initiative. These include the following:

- Techniques and tools for semantic data organisation in a mixed environment (cloud + local storage) linked by their semantic “meaning”, especially considering human values like fairness and adhering to GDPR requirements and taking into account advances in low power-based systems e.g. IoT devices;

(it doesn't mind if you make a typo in your research, or if you write just a single word like 'fish', Google knows if you are attempting to look for a golden fish for your bowl or if you are looking for a 4-stars restaurant!)

- Remote / Automated enforcement of data handling rules; using AI to extract semantic meaning and innovative solutions to contextually manage data release;
- Ensuring correctness and authenticity of entire value chain verification mechanisms to ensure accuracy;
- Would blockchain technologies be feasible for next generation systems based on low power;
- Distributed Ledger Technologies for open platforms;
- Advanced semantic search technologies;
- Specialised Search technologies and tools (e.g. IoT, Big Data, AI, ...);
- Supervisor systems (e.g. AI supported) for tackling online disinformation;
- Work is needed to address the human / social effect of living in an era of sensationalism and receiving regular online disinformation;
- Innovative uses of blockchain technologies to support traceability of data and information.

These R&I topics will feed into the NGI Consultation platform and Knowledge base <https://consultation.ngi.eu/> of the SpeakNGI.eu project and to the RIA on Search and Discovery (expected to start in November, 2018), which will fund innovative and agile projects in three dedicated NGI areas: Privacy and Trust Enhancing Technologies, Decentralised Data Governance, and Discovery and Identification technologies.

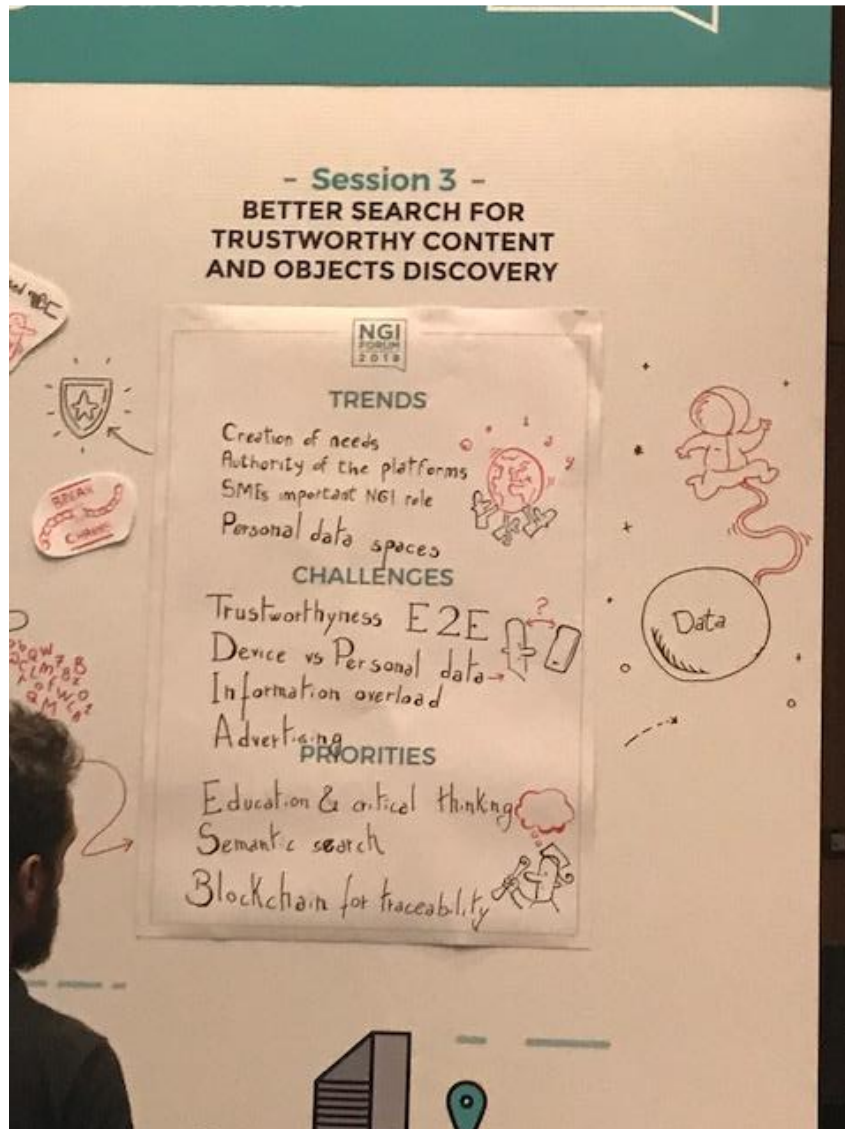


Figure 7: Bulletin board summary presented during the final plenary session of NGI Forum, Porto, 13th Sept. 2018.