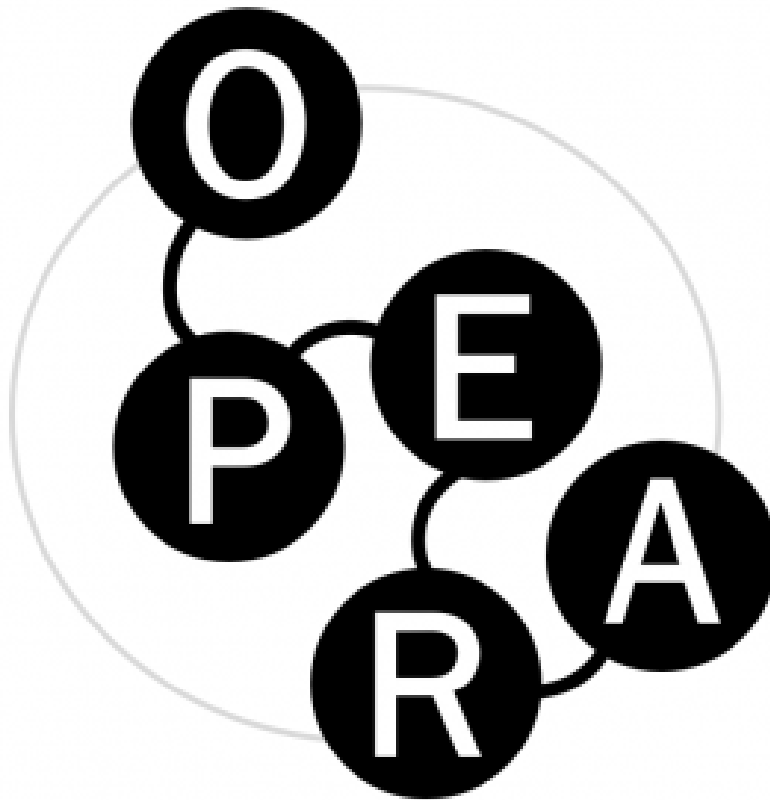# ONTOLOGIES AND TOOLS FOR IMPACT DATA COLLECTION, ANALYSIS AND VISUALIZATION

OPERA Project Report no. 2

18 November 2020

Contributor credits

Specified using the CRediT Taxonomy contributor roles:

Conceptualization
Ideas; formulation or evolution of overarching research goals and aims

Methodology
Development or design of methodology; creation of models

Software
Programming, software development; designing computer programs; implementation of the computer code and supporting algorithms; testing of existing code components

Validation
Verification, whether as a part of the activity or separate, of the overall replication/reproducibility of results/experiments and other research outputs

Formal Analysis
Application of statistical, mathematical, computational, or other formal techniques to analyze or synthesize study data

Investigation
Conducting a research and investigation process, specifically performing the experiments, or data/evidence collection

Resources
Provision of study materials, reagents, materials, patients, laboratory samples, animals, instrumentation, computing resources, or other analysis tools

Data Curation
Management activities to annotate (produce metadata), scrub data and maintain research data (including software code, where it is necessary for interpreting the data itself) for initial use and later reuse

Writing – Original Draft
Preparation, creation and/or presentation of the published work, specifically writing the initial draft (including substantive translation)

Writing – Review & Editing
Preparation, creation and/or presentation of the published work by those from the original research group, specifically critical review, commentary or revision – including pre- or postpublication stages

Visualization
Preparation, creation and/or presentation of the published work, specifically visualization/data presentation

Supervision
Oversight and leadership responsibility for the research activity planning and execution, including mentorship external to the core team

Project Administration
Management and coordination responsibility for the research activity planning and execution

Funding Acquisition
Acquisition of the financial support for the project leading to this publication.

AUTHORS

David Budtz Pedersen – ORCID ID 0000-0001-7861-7068 – email: davidp@hum.aau.dk

ROLES: Conceptualization, Methodology, Investigation, Writing Original Draft, Writing, Review & Editing, Project Administration

AFFILIATION: Aalborg University

Pelle Annfeldt Israelsson – ORCID - email

ROLES: Writing Original Draft, Resources.

FORMER AFFILIATION: Aalborg University

TABLE OF CONTENTS

1. INTRODUCTION

This review seeks to map relevant ontologies, tools and platforms designed to track, capture and visualize research impact. Looking at existing tools and platforms, the review seeks to create a better understanding of how different academic activities are represented and how (meta) data is collected, presented and connected to research funding agencies and knowledge organisations. Through a descriptive analysis, the report describes a number of methods and applications relevant for tracking broader impact activities and outputs of Open Science with a special focus on humanities and social sciences.

The platforms and tools chosen for this review is a small representation of a larger field in which data-driven impact analytics are rapidly emerging. The examined tools and platforms are developed by a broad range of actors, including universities, companies and libraries. The examined tools and platforms have been selected in order to present a balanced table of hands-on approaches. In preparation of the review, the authors have used and tested the tools and a desktop study has been carried out with special focus on applicability of tools and ontologies for the DEFF OPERA VIVO. The tools described have been teste during the core period of the DEFF OPERA project (2018-2020), and relevant literature relating to the tools and platforms have been examined. In addition, we have arranged meetings and interviews with key actors from the tools and platforms under examination.

2. TOOLS FOR COLLECTING RESEARCH IMPACT DATA

2.1. ELSEVIER PURE

Pure is an enterprise research management solution that aggregates an organization's research information from numerous internal and external sources into a single platform. PURE was born out of a collaboration between seven Danish universities (AUB, DVJB, RUB, AU, SDU, KU and ASB), and has been tailored to be a research registration system and e-archive platform for Danish university and research environments. It was acquired by Elsevier in 2012, and is today one of the world's leading research information management system with 250 implementations worldwide containing profiles for over 185,000 researchers.

Pure can be configured to satisfy a number of organizational requirements. It is used to capture data about research, teaching and third-mission activities; it is used to validate bibliometric data that feed into national bibliometric scoreboards; it is used to profile researchers to the public by

aggregating and presenting publicly available data about research, teaching and third-mission activities; it is used to identify experts and track funding and projects, and finally the platforms allow for a high number of different options to report, analyse, track, monitor, visualize and showcase multiple aspects of universities' performance using various plug-ins.

Sample of content types in Pure:

Person	Awarded grants	Journals
Org units	Projects	Prizes won
Equipment	Activities	Press clippings
Funding opportunities	Research outputs	Impacts
Applications	Datasets	Courses Taught

Pure can search external data sources to identify outputs associated with researchers and institutions. It retrieves metadata automatically from e.g. Scopus, Web of Sciences, PubMed, Embase, Mendeley, arXiv, Worldcat, CrossRef, JournalTOC, CAB Abstracts, SAO/NASA Astrophysics Data System (or from files e.g. BibTeX or RIS). Pure can be used to create, update, and correct researcher profiles by retrieving publication lists from relevant internal and external databases. Also, among the many subscription options, it is possible for clients to add additional content using a wide range of templates that provide structure for reporting e.g. adding information about scholarly activities, funding, intellectual property, creative works and "impact modules" that can reveal a more comprehensive picture of the organisation's expertise and become more discoverable by potential collaborators.

Pure's Basic Activity Typology

| |
|---|
| Attending an event |
| Conference organisation or participation |
| Organisation or participation in workshops, courses, or seminars |
| Public/private sector consultancy and other employment |
| Employment with any other public or private company including your own company |
| Consultancy |
| Public Sector Consultancy |
| Starting your own company |
| Editorial work and peer review |

Peer review of manuscripts
Journal editor
Series editor
Editor of unfinished research anthology/collection
Examination
External examination
Internal examination
Hosting a guest lecturer
Hosting a guest lecturer
Memberships
Board duties in companies, associations, or public organisations
Membership of committees, commissions, boards, councils, associations, organisations, or similar
Membership of research networks or expert groups
Membership of review committee
Talks and presentations
Guest lecturers
Conference presentations
Talks and presentations in private or public companies
External teaching and course activities at other universities
Visiting another research institution

A number of the above categories can be used to track and capture research impact and third-mission-activities. Where is Pure is focused on registering outputs and activities, registrations do not indicate or document end-effects or real-world change of behaviour, practices or policies. However, data from Pure can be used to create self-assessment, data for external review panels, and case studies that can be assessed during peer panel reviews or by submitting impact data to public and private funding agencies.

In addition to the above basic activity typology, Pure has created an improved data protocal to help organisations demonstrate and report on research impact. Designed in collaboration with the Pure User Group and drawing on lessons learned during the UK Research Excellence Framework 2014 (REF2014), the impact data model allows universities to capture both narrative and background structure information. This includes relations to publications or research projects, or activities such as conference presentations or media appearances. Pure continues to

develop the impact module for targeted custom groups e.g. completion and delivery of the REF 2021 module for UK clients. This module is specifically designed for UK customers to be able to select content that meets the needs of the REF assessment, combining linked data across the platform to REF impact case studies (submitted the Research England). A more rudimentary format has been released for Danish universities:
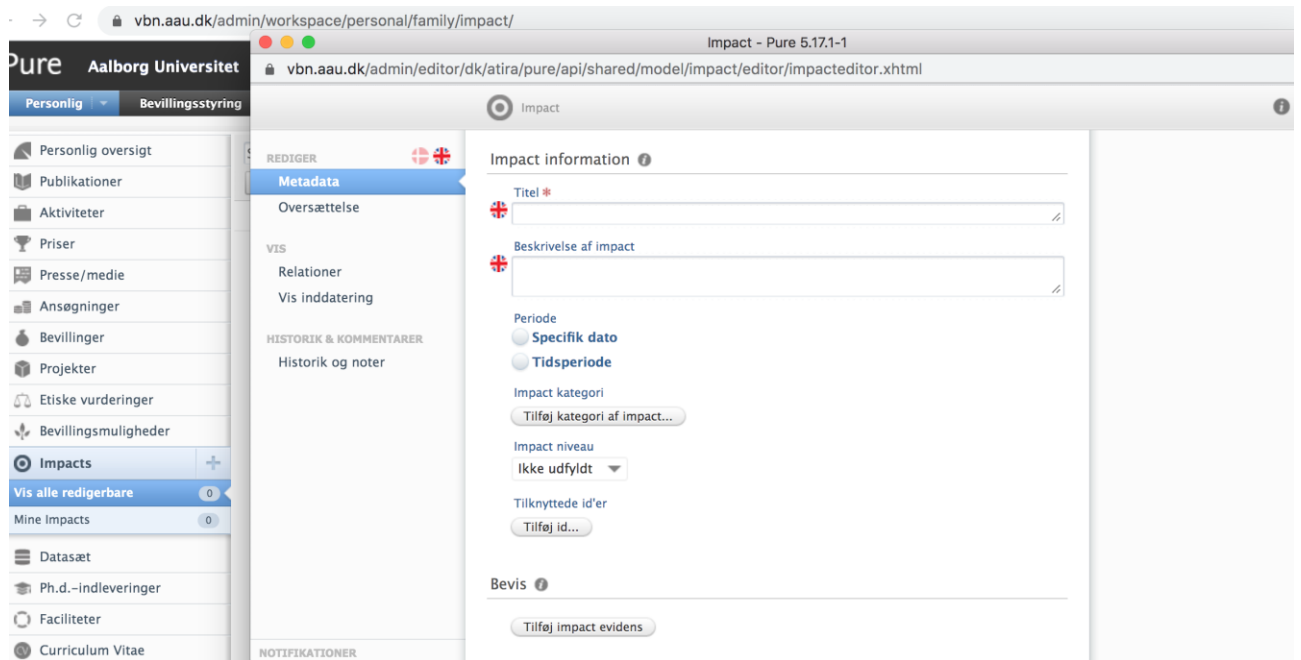
The strength of Pure is the platform's ability to showcase key performance and research expertise, using the Elsevier Fingerprinting technology to create research expertise pages that help to showcase the extent, through the range of connected content that highlight all of the researchers, achievements, projects and impact that institution has had on a specific field.

Athena University

Help & FAQ

Home   Experts   Research Units   Research Output   Projects   Datasets   Activities   Equipment

Enter search terms...

Chris Robertson, M.Sc.CS, PhD
Department of Statistics & Modelling Science

62 Citations   **7** h-Index

Phone
1 (212) 539 5555
1101 South Bancroft Street
New York NY 00000
USA

E-mail
chris@athenauniversity.edu

Website
www.elsevier.com

2009 ▮▮▮ 2016

👤 Overview   🌀 Fingerprint   🔖 Projects (5)   📓 Research Output (26)   ☰ Datasets (3)   💼 Activities (3)   👤 Similar Experts (10)   •••

🌀 Fingerprint

The Elsevier Fingerprint Engine mines the text of the experts scientific documents - publication abstracts, awards, project summaries, patents, and other sources to create an index of weighted terms which defines the text, known as a Fingerprint. By aggregating and comparing Fingerprints, the Elsevier Fingerprint Engine enables users to look beyond metadata and expose valuable connections among people, research units, publications, and ideas.

Sort by | Weight | Alphabetically

Medicine & Life Sciences

Vaccines   Population   Women   Infection   Disease   Methods   Influenza, Human   Vaccination   Alcohols   Risk

Confidence Intervals   Humans   Prevalence   Fibrosis   Theoretical Models   Databases   Hospitals   Streptococcus pneumoniae

Hepacivirus   Health   Reproduction   Time   Therapeutics   Child   Urine   Schools   Death   Disease Outbreaks

Laboratories   Data Collection   Residence Characteristics   Seasons   Influenza Vaccines   Soil   Fenretinide   Family Characteristics

Motor Activity   Infertility   Diagnosis   Alcohol Drinking   Patients   Tamoxifen   Papillomavirus Vaccines   Hospitalization

Hepatitis C   Blood   Referral and Consultation   Micronutrients   Malawi   Drug Users   Comprehension   Oseltamivir   HIV

Pandemics   Epilepsy   Propensity Score   Immunization   Breast Neoplasms   Pneumonia   Bacteria   Pilot Projects

Magnetic Resonance Spectroscopy   Maternal Welfare   Pediatrics   Conjugate Vaccines   Pneumococcal Vaccines   Immunization Schedule

Cities   Questionnaires   Mass Spectrometry   Evaluation Studies   Public Health   Health Status   Probability   Pharmaceutical Preparations

Placebos   Insulin-Like Growth Factor I   Primary Health Care   Human papillomavirus 18   Liver Diseases   Cerebrospinal Fluid   Cell Biology

Morbidity   Female   Metabolomics   Serum   Work   Biological Markers   Human papillomavirus 16   Length of Stay   Men

Hospital Records   Administrative Personnel   Delivery of Health Care   Logistic Models   Breast   Persons   Male   Travel   Virology

Age Groups

Earth & Environmental Sciences

Materials, Products and Resources

smoke

Biological Phenomena

mortality

In sum, Pure enables subscribers to exhibit their organization externally, promote researchers, outputs and networks and, and facilitate collaboration via a specialized web portal, known as the Pure Portal. See more: https://www.elsevier.com/solutions/pure

2.2 RESEARCHFISH

Researchfish is a monitoring tool consisting of an online system which collects data from researchers about their outputs and progress – often related to monitoring and reporting on specific grants. The system is implemented by more than 150 organisations world-wide with 100.000 individual users, including UK Medical Research Council, Cancer Research UK, Gates Foundation, Wellcome Trust, Novo Nordisk Foundation, Lundbeck Foundation, Velux Foundations as well as Oxford University, University of Cambridge, and Alberta University to name a few. Researchfish enables research funding agencies to track the progress and output of their recipients in real-time, by collecting input data about from researchers regarding their projects. Documenting and collecting data about the impact of grants has significant interest among funding agencies. Funding agencies are under increased scrutiny to demonstrate their accountability to society, and are increasingly applying an evidence-based methodology to collect structured and linked data about scholarly output and impacts. From the perspective of funders, enriched metadata provides an opportunity to refocus their strategy, portfolio and instruments. The Novo Nordisk foundation have stated in their allocation strategy that their research should contribute positively to education, quality oflife, improved treatment and have a high micro- and macroeconomic impact in society. Researchfish allows funders to track these performance goals by capturing a wide range of impact activities and outputs.

Researchfish Basic Activity Typology

Publications
Collaborations
Further Funding
Next Destination & Skills
Engagement Activities
Influence on Policy
Research Tools & Methods
Research Databases & Models
Intellectual Property & Licensing
Medical Products, Interventions and Clinical Trials
Artistic & Creative Products
Software & Technical Products
Spin Outs
Awards and Recognition

| Other Outputs & Knowledge / Future Steps |
| Use of Facilities & Resources |

For full list and subcategories, see [link](#).

Researchfish is used as an instrument for:

- Monitoring whether the funding allocation portfolio fulfils the social objectives of the funder (vis-à-vis strategy, mission, mandate) and to ensure an evidence-based approach to decision-making on distributions and adjustment of funding instruments and activities.

- Support the ongoing improvement and refinement of research management, priorities and resource allocation based on the clients' results and impact.

- To create evidence-based communication and reporting fund performance results, and to create a picture of the impact of grants on education and innovation activities in the surrounding community.

In practice Researchfish serves as a more advanced (and in principle) less burdensome interim, annual and final reporting platform for grant holders. In addition to submitting interim and final reports, the tool collects data throughout the project lifecycle based on self-reporting and validation by the funding agency, thus making data more readily available, comparable and useable. Researchfish was originally designed to document the effects of health science and pharmaceutical research, but is currently used in a wide range of other disciplines, including science, social sciences and human sciences. There is ongoing debate about the potential of the instrument and possible implementation in other areas (for further analysis of strengths and weaknesses of using Reseachfish in impact assessment, see WP2a Report).

The platform makes it possible to zoom in on specific types of impact, such as the impact of funded research on policy-making. Using this template, funders are able to capture details of any significant impact on policy or practice that has been realised as a result of funded research. Research that has impacted on policy areas that effect health, society and the economy, such as housing, transport and schools should also be reported.

Sample of policy impact indicators in Researchfish:

Implementation circular/rapid advice/letter to e.g. Ministry of Health
Influenced training of practitioners or researchers
Citation in clinical guidelines
Citation in clinical reviews
Citation in other policy documents
Citation in systematic reviews
Membership of a guidance committee
Participation in a national consultation
Participation in advisory committee
Gave evidence to a government review

Researchfish has interoperability with Pure and can retrieve metadata automatically from e.g. Scopus, Web of Sciences, PubMed, etc. (limited to publications with DOIs). Researchfish works by bulk upload capability as well as manual entries for books, book chapters, monographs, policy briefings and reports, journal articles, conference proceedings and abstracts etc. Data can be exported in various standard formats and used for graph designs and visualisations (external to user interface).

2.3 REACT VIVO

For research purposes, the ReACT research team at Aalborg University has developed an impact registration system and taxonomy operationalised in VIVO. VIVO is a web-based, open-source suite of computer software for managing data about researchers, scientists, and faculty members. VIVO uses semantic web techniques to represent scholars and their work (Ding et al. 2011). The VIVO basic ontology incorporates elements of several established ontologies, which are used to describe a number of different roles of faculty members, including research, teaching, and third mission activities. During the research project "Responsible Impact" (2016-2020), hosted by Aalborg University and funded by the Obel Family Foundation, the research team constructed a new domain taxonomy for "research impact" with a division of entities and activities distributed among 100+ categories. The ontology was implemented in VIVO. While suspending several of the existing layers of the basic ontology and adding a new architecture for tracing and capturing the diversity of impact activities, the ReACT VIVO utilizes several of the platform's existing features for publication records, geolocation, and linked data.

The domain taxonomy for research impact was constructed during a multi-level research exercise comprising: step 1) a review of existing indicators; step 2) a participatory mapping of research impact based on input from AAU scholars, refined by iterative staff involvement, and step 3) a close-reading of curriculum vitae stemming from a group of "prototypical" scholars with special attention to their third mission activities. These four data sources provided the input for constructing a web of NN impact categories distributed among NN sub-categories.

Impact entities were further separated into two disjoint categories of *input* signals and *output* signals, the former comprising for example invitations, nominations and publicity, the latter comprising products, processes and outputs conceived as leading to impacts (i.e. impact pathways). Output signals were then separated into another set of disjoint categories: impact signals based on *participation* in-person (events, projects, activities, performances) and impact signals based on artefacts and *products* (created, authored, curated, externalized) by the unit of assessment.

Sample of content types in ReACT VIVO

Impact pathway	Categories	Sub-categories
Participation in Person		
	Academic Event	
		Conference
		Seminar
		Workshop
		Meeting
		PhD Defence
		Presentation
	Non-academic Event	
		Meeting
		Presentation
		Workshop
		Open Days

		Celebration
	Interactive research	
		Meeting
		Workshop
		Presentation
		Observation
		Experiment
		Intervention
	Formal engagement	
		Consulting
		Start-up
		Science Advice
		Board member
		Working Group
	Teaching	
		Pre-school
		Primary school
		Secondary school
		Higher education
		Professional training
		Life-long learning
		Open university
	Quality assurance	
		Examination
		Employment assessment

		Endorsement
		Editorial work
		Peer review
		Curation
		Evaluation
		Supervision
	Media appearance	
		Interview
		Content creator
		Opinion
		Criticism
Participation by product		
	Media	
		TV
		Radio
		Newspaper
		Magazine
		Book
		Edited website
		Unedited website
		Documentary
		Blog
		Social media
	Aesthetic products	
		Artwork
		Movie

		Theatrical play
		Exhibition
		Festival
		Performance
		Book (catalogue)
	Academic outputs	
		Journal
		Monograph
		Edited vol.
		Proceedings
		Professional magazine
		Newsletter
		Website (research)
		Curricula
		Database
	Policy outcomes	
		Green paper
		White paper
		Policy docs
		Law texts
	Commercial output	
		Patents
		License
		Utility model
		Proof-of-concept
		Design

		Trademark

The domain taxonomy was operationalised in VIVO by developing two interfaces: an interface for registration of impact activities distributed among the categories introduced above, and an interface for (selected) public presentation of performance and scholarly activities. In the course of a 12-month period spanning 2018, 100 active researchers from Department of Communication and Psychology at Aalborg University was sampled and invited to participate in data collection. 45 researchers signed up and participated in extensive data collection, which was compensated for with 12 hours per person by the university. In addition to the workload (estimated to 12 hours in 12 months), an 'impact manager' was appointed by the research team to facilitate and support data collection. For some participants in the test population, the impact manager collected almost all impact data, while other participants used the registration interface to supply real-time data on a daily, weekly, or monthly basis.

After finalising data collection, analysis, integration and interpretation was carried out. The strength of using VIVO is the extended use of linked data, which means that many of the impact activities registered in the database can be linked, thus leading to a more comprehensive picture of the different networks of impact activities. For validation purposes, the data for a selected subset of the test population has been integrated into five impact case studies. Case studies have been selected in order to display different styles of research and impact profiles. Common for all five case studies is the commitment to use VIVO impact data, bibliometric data (BFI) and altmetric data (social media and Infomedia) to establish a comprehensive data environment as the baseline for constructing the case study (*forthcoming* 2020).

Beyond recording real-time research activities on the VIVO platform, the project is experimenting with novel ways of representing and visualising outcomes, e.g. by deploying techniques from infographics and graph databases. The resulting aggregated data are featured in the five case studies, enriched by qualitative interviews with faculty members and further augmented by bibliometric records and altmetric profiles. Together, VIVO impact data, bibliometric data and case studies create a *portfolio* for the involved research groups that can designed to support career development, profiling research, and provide visibility and information about expertise and performance (for further description of methodology, see WP2a Review).

ReAct aims to develop a model of impact assessment, which is based on the number of relations, networks and connectivity with actors outside and beyond academia. By explicitly drawing on a *responsible* approach to research assessment, ReAct takes into account the diversity of impacts by creating a participatory taxonomy that allows researchers to create their own impact typology and redefine existing categories. By using the notion of "productive interactions" as the baseline

for tracing the impact of OS activities, it is possible to explore a wide range of collaborative activities, such as policy interactions, citizen interactions, industry interactions and media interactions (Spaapen & Drooge 2011; Budtz Pedersen et al. 2020). See more: react.aau.dk

3. SELECTED TOOLS FOR ANALYSING IMPACT DATA

3.1   DIMENSION

Dimensions aims to provide a data-rich environment which allows for the development and dissemination of metrics, developed for the research community. The Dimensions platform is intended as a tool for integration research information, performance, data and collaborations. Dimension is a part of Digital Sciences. One of the aims of Dimensions is to enable the research community to develop and implement metrics to solve specific use cases. Dimensions seeks to undertake this role as a data and dissemination platform for these efforts.

Dimension have collected its data from publishers using DOI's as identifiers, and have a cover rate of 97-99 per cent of publications form DEFF OPERA member: KU DTU in OPERA's test (OPERA, 2019). In contrast, Microsoft Academic and Google Scholar also index web content from crawlers. This means that the citation counts in Dimensions are lower than those of Microsoft Academic and Google Scholar but has the important advantage that its data is less easy to contaminate. Generating fake papers with self-citations and posting them to academic domains is an effective way to spam indexes that look for academic content online (Thelwall, 2018).

3.2   ALTMETRIC

Altmetric is an alternative bibliometrics platform belonging to Digital Sciences. The company has developed and sophisticated alternative indicators for research impact, such as social media presence and online references. Alternative metrics are used to indicate how scientific publications are cited and mentioned other outside the academic literature. Altmetrics are measured by counting the times articles, dataset, reports etc. are referenced on internet platforms. Social media reference to articles is often shown next to citations in journal websites, introducing this new measuring into academic metrics are dominated by Altmetric.com and their donut badge. In order to generate its score, Altmetric is crawling different social media and academic dissemination platforms; Twitter, LinkedIn, Google+, Sin Weibo, YouTube, Reddit, blogs, Wikipedia, Mendeley, patens, mainstream media and public policy documents. Scores are generated by searching for a series of identifiers; PubMedID, arXiv ID, ADS ID, SSRN ID, RePEC ID, Handle.net identifiers URN, ISBN, DOIs across platforms looking for reference to academic work.

Altmetrics claims that all direct reference is tracked. However, second order citation, where a blog post or news stories reference a scholarly output are not tracked.

Reference and scholarly platforms have introducing these new standards in their user interface (UI) with little to no consideration if these metrics are valid, or demonstrates how research activity are perceived in the scientific community. Some test results have shown that altmetrics are susceptible to twitter-bots and manipulation (Adie, 2019). Furthermore, the index is un-weighted and all references are counted as equal, implying all citations are the same; a mention in a policy document and on a blog are thereby counted as the same. The donut changes colour depending on the source of the citation, leaving it to the reader to determine if the citation carries importance.

Applying alternative metrics to a VIVO platform could be beneficial as this metric have become a staple indicator on most publication platforms. The metric provides some understanding of how the research community is perceiving the work, and if any other communities are reacting and talking about the research. This metrics are available as add-on in the interface, making it a relatively easy addition to the UI. However, no reward system is established at the university or funding level to provide incentives for researchers to interact on social media.

3.3   PLUM X

Plum X became part of Elsevier in 2017. Plum X is driven by the observation that "telling the story of research is becoming an increasingly important part of understanding the impact of that research" (McEvoy, 2015). Traditionally, citation counts have been used to classify a study's impact but this type of reporting is hindered by the time-lag that often accompanies scholarly publications as well as other issues. Plum Analytics uses real-time information to provide insights into the attention attributed to a specific piece of research that might be beneficial for customers who need to track and understand how research output is being received (funders, universities etc). Plum X argues that this type of metric can be used for reporting and evaluation purposes, to help securing funding or for other reasons.

Like Altmetric, Plum X is based on a broader interest in the uptake of research than can be expressed by citations alone. Data is collected from the places, scholars and stakeholders interact with research artefacts such as articles, clinical trials, blog posts, grants, books, theses/dissertations, webpages and more. These metrics are then categorized into Usage, Captures, Mentions, Social Media and Citations. PlumX integrates research metrics for all types of scholarly research output. It categorizes metrics into five groups:

- **Usage** – a way to signal if anyone is reading the articles or otherwise using the research. Usage is the number one statistic researchers want to know after citations. Examples include clicks, downloads, views, library holdings and video plays.

- **Captures** – indicate that someone wants to come back to the work. Captures can be a leading indicator of future citations. Examples include: bookmarks, code forks, favorites, readers and watchers.

- **Mentions** – measure activities such as news articles or blog posts about research. They are a way to tell that people are truly engaging with the research. Examples include blog posts, comments, reviews, Wikipedia links and news media.

- **Social media** – includes tweets, Facebook likes, etc., that reference the research. Social media can help measure "buzz" and attention; it can also be a good measure of how well a piece of research has been promoted. likes, shares and tweets.

- **Citations** – include both traditional citation indexes such as Scopus as well as citations that help indicate societal impact, such as clinical or policy citations. Examples include citation indexes, patent citations, clinical citations and policy citations.

3.4   COBALTMETRICS

Cobaltmetrics collects alternative bibliometric indicators by crawling the web to index hyperlinks and identifiers. Crawling and analysing a wide range of webpages searching for references, the platform collects links between documents and researchers (Boruta, 2019). Cobaltmetrics applies a wider search engine than backlink databases and altmetrics aggregators, not only searching social media and webpages for identifiers reference to publications and researchers. The metric is generated by mining data across 180 languages and 175 shortneres, crawling publications, books, clinical trials, patents, software artifacts, derivative works (Cobaltmetrics.com). Key principle of Cobaltmetrics "*it is not up to bibliometrics aggregators to decide what is citable*" (Boruta, 2019), applying a broad web search where all reference are counted and then letting the end users use the identifiers that they are most comfortable with, letting Cobaltmetrics preform the heavy computational lifting. Cobaltmetrics use of web crawlers and search tools, makes the results more comprehensive reaching a larger set of citation, links and mentions. However, the UI and the results of searches is still being refined and tested. When testing Cobaltmetrics on DEFF OPERA members the results was not impressive.

3.5   GRID

GRID is part of Digital Science. The platform has been developed as an open database of research institution identifiers which enables users to enhance the data by allowing all research institutes, researchers, articles and funding to be directed to the right institutional labels. Thereby minimising the work required when linking articles and grants together with researchers and university using a unique and persistent identifier. The result is the most clean and accurate data for robust reporting. Grid provides a platform from where information about education institutions, companies, healthcare facilities and hospitals, and government institutions are represented. Grid also adds a geolocation and other known identifiers (Crossref, Wiki, URL and OrgRef) to the institutions, making Grid an opportune platform when referencing institutions.

The introduction of geo-coordinates and affiliation metadata will be beneficial when trying to understanding how collaboration between institutes are created. Looking at how geographical distance could be an indicator for how collaboration between researchers are created, or which grants are sought. The platform provides only metadata to institutions and should be used as a platform which can enhance existing data.

Metadata:

ID	grid.5117.2
Types	Education
Established	1974 CE

External links:

Institute Links	http://www.en.aau.dk/
Wikipedia	http://en.wikipedia.org/wiki/Aalborg_Un...
isni	0000 0001 0742 471X
Crossref	501100002702
OrgRef	382707*, 33784804
	Q601956
	* Preferred ID

Alternate Labels:

Aliases	Aalborg Universitet
Acronyms	AAU
Danish	Aalborg Universitet

Relationships:

Related Institutes	Aalborg Hospital

Aalborg - Denmark

REPORT AN ERROR

GeoNames

Type	Name	GeoNames Code	GeoNames ID
City	Aalborg		2624886
Admin 2 Region	Ålborg Kommune	DK.19.851	6543934
Admin 1 Region	North Denmark	DK.19	6418540
Country	Denmark	DK	2623032

NUTS

Level	Name	NUTS code
NUTS 3	Nordjylland	DK050
NUTS 2	Nordjylland	DK05
NUTS 1	DANMARK	DK0

4. VISUALISATION TOOLS AND GRAPH DATABASES

4.1   GEPHI

The visualising tool Gephi is an open source visualization software tool which is used for network visualizations of data. Its main quality is the software's ability to spatialize a scattered network into a map by applying an algorithm on the data. The default, and most commonly used, algorithm is ForceAtlas2 (Jacomy, Venturini, Heymann, & Bastian, 2014). The fundamentals of the ForceAtlas2 are straightforward: when the algorithm runs, the nodes repulse each other, where the edges are used as a way of attracting the nodes, the more edges that connects two nodes the stronger the pull becomes, this movement is generated by a mathematical interpretation of the data. The algorithm has been designed with simplicity in mind. The accessibility of using the program have also been in consideration doing the design, where a trial-error approach has been used to improve the learning curve of the user (Jacomy et al., 2014)

Data landscape & mapping

Using Gephi's mapping tools creates a map where different nodes are placed by their connection to each other. This function helps force the different nodes together or apart, thereby reveal regions where there is a higher density of nodes and regions where there are almost no nodes. This way, placing of the nodes at the map are determined by the nodes' mathematical properties. When creating a map by using Gephi, one has created a data landscape, which can help to navigate the data. The data landscape shares some attributes with a geographical map. (Jacomy, pers. comm.)

Unlike a geographical map, the data landscape is not arranged after a predetermined set of values on the x and y axis, where the different landmarks are placed on the map correlated to the distance between the two landmarks. The data landscape only uses the connection between two nodes to determine where the nodes are placed in correlation to each other, therefore the nodes in the data landscape do not have a set of values on the x and y axis, this allows for the data landscape to become more fluid because the nodes place in the map is only determined by the connection to other nodes, therefore the data landscape have the properties of being flip without losing its validation.

If one were to only use the mathematical properties of the nodes and divide the nodes into different tables, the fuzziness of the map would be lost. By using visual tools, the regions where the nodes a placed become more blurred and therefore allows for a negotiation of how to read the different regions of the map. If one looks at a region of nodes one can have a problem determine where the boundaries of the region begins, thereby allowing nodes that a mathematical division otherwise would have not included and thereby giving a more nuanced map. Also, when looking at a data landscape there are areas where there a fewer nodes or none at all. These areas are called 'structural holes', and when looking at the map these structural holes help identify the different regions by highlighting where there are empty zones in the map, which thereby gives an indication of where the data is more clusters.

Clusters

When looking at regions one can identify smaller clusters within these regions. These areas of the map can be identified either by their connectedness or the size. When looking at clusters where nodes shares a lot of connections the cluster have a higher density, this can indicate a community where a lot of activity is taking place. Different high-density clusters exist 1) there is a central placed node where most of the surrounding nodes connect through this node. 2) the different nodes in the cluster circulate a lot of information without a single central node, and therefore create a cluster with a high density but no central node. Gephi can be used as a mapping tool for a series of different data types and formats, making it possible to visualise co-authorship, twitter mention, collaboration and etc. Gephi is an opensource tool where the users are provided a platform which are highly changeable and fluid. The visualisation of data is useful when large and complex dataset is presented, by opening the data to easy interpretation.

Figur 2: Screenshot from Gephi

4.2 VOSVIEWER

VOSviewer is a mapping tool for constructing visualised bibliometric networks, developed by CWTS at Leiden University. VOSviewer is constructed by the same principle of data landscape and mapping as Gephi, there is a difference between Gephi and VOSviewer in their cluster generating. VOSviewer is constructed as a bibliometric network visualising tool, being specialised with academic publication data. Designed to operate with input from Web of Science and Scopus data, optimised towards these data sources. A natural language processing function enables VOSviewer to created maps based on term co-occurrence form large text-based corpuses. VOSviewer uses a distance-based algorithm (force vector) to arrange the different nodes in relation to each other.

VOSviewer are tailored to work with larger maps, using the Lauvain algorithm to manage modularity (colour) on large network (Waltman & Van Eck, 2013). In the UI of VOSviewer the nodes are indicated by circles and sized by the occurrence in the dataset, intuitive giving the user an understanding of significant. Furthermore, the labels are scaled by node size, meaning smaller

nodes do not have visualised labels, to avoid overlapping labels, only a subset of all labels is displayed, until a closer examination of the map is conducted (van Eck & Waltman, 2010).

VOSviewer is by design optimised to support and visualise bibliometric networks, making the tool well suited for visualising researches publishing output. The natural language function makes it possible to visualise the written output from researchers, illuminating their position in the scientific community, based on their publication by analysing their use of terms as a identifier for topic interest.

Figur 3: VOSviewer

4.3   CITNETEXPLORER

CitNetExplorer is a software tool for analysing citation networks of scientific publication created by CWTS from Leiden University. The tool uncovers and illuminates the citation pathways networks. The network is created by importing citation lists form journals and books published during a period of time, the citations are then arranged by article who cites (source) and cited article (target), connected by citation (directed edges)(Van Eck & Waltman, 2014). The defining asset of this tool is the timeline, most other network visualisations tools (VOSviewer, Gephi etc.) are arranged by the edges, using a force vector algorithm to create the visualised network. CitNetExplorer differ from other visualising citation network tools by its extensive function for exploring and analysing citation network, and functionality for drilling down into citation network

(Van Eck & Waltman, 2014). The functionality of drilling down, are used by CiteNetExplorer to transits from full citation network consistent of several million publications, down into subset of manageable size, these subsets are all dealing with a specific topic of interest. Making the tool relevant for reviews, literature searches, mapping of scientific disciplines and systematic literature retrieval. The searches can be based on author, journal, title, ect (Van Eck & Waltman, 2014).

The timeline function of CiteNetExplorer arranges the articles in a chronological order, the vertical dimension indicating the oldest at the top and the newest at the bottom (see Figur x). The horizontal dimension the publications are arrange by their connectedness, publication are located close to each other if they are closely connected in the citation network (Van Eck & Waltman, 2014). This arranging of the articles by year and citation allows researcher to follow the direction of ideas and concepts over time and publications. CitNetExplorer uses bibliometrics to showcase how concepts and ideas travel through publications, by visualising the citation partway. Using citations as indicator for origins of ideas a is a classic academic discipline, used in literature- and systematic reviews. CitNetExplorer allows researchers from all disciplines to visualise citation pathways early in the research circle, as the tool allows researchers an overview of the topic field.
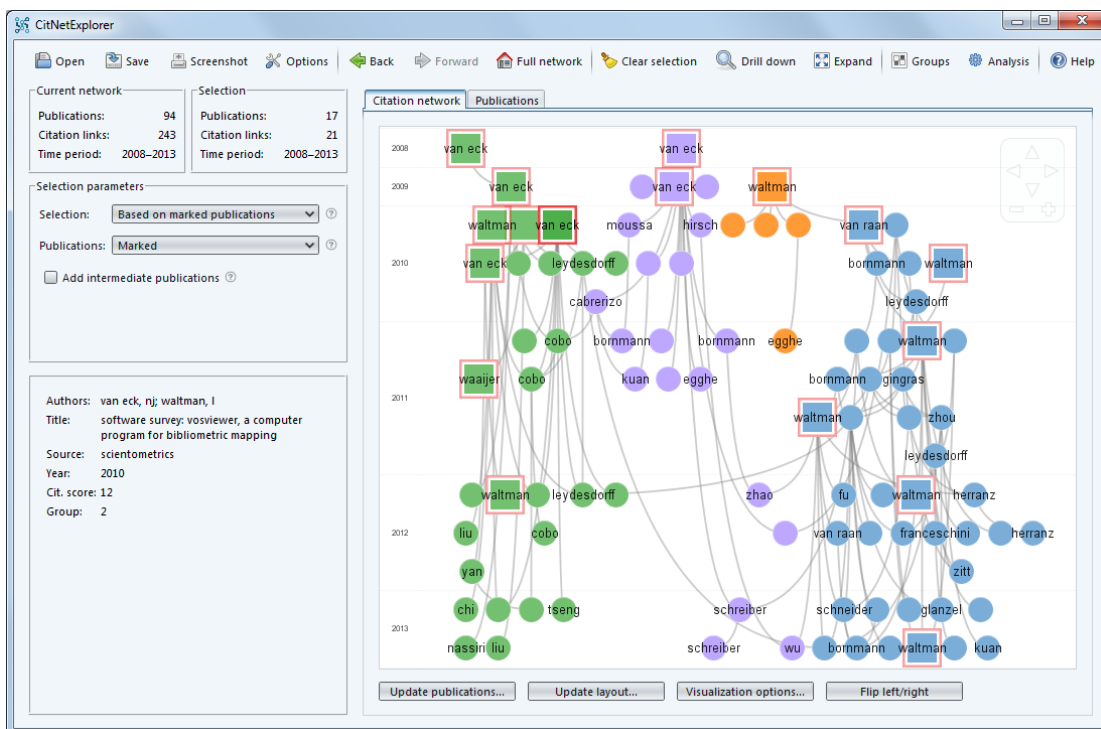

Figur 4: CitNetExplorer

4.4   RAWGRAPHS

RawGraphs is an open source collection of data visualisation frameworks developed by DensityDesign Research lab at Polytechnic University of Milan. RawGraphs aim at providing a missing link between spreadsheet applications, making the visual representation of complex data accessible. The visualisation are created by having the user input structured datasets (CSV, TSV, XLSX and JSON), the structuring of these dataset and the values (date, numbers, string and labels) determines which visualisation charts can be applied to the dataset (Mauri, Elli, Caviglia, Uboldi, & Azzi, 2017). The RawGraphs output is SVG, PNG and JSON theses formats are chosen as they represent different aspects of usage. The SVG format provides a geometric description of the different elements in the visualisation, making the format available for designers to manipulate, in order to create the correct representation of the data in presentations. PNG is a format which are easily available for embedding into presentation and sharing online. The JSON format, is a text based visualisation, and not a graphical format, it is useful when dynamically recreating the visualisation (Mauri et al., 2017).

RawGraphs pallet of different visualisations can be used as a starting point from where research related actions taken by researchers in SSH can be presented. As a method to communicate how researcher interact with the surrounding society. By visualising dataset, narrative anecdotes can travel and become empiric and tangible representations of how researchers in SSH interact with society.
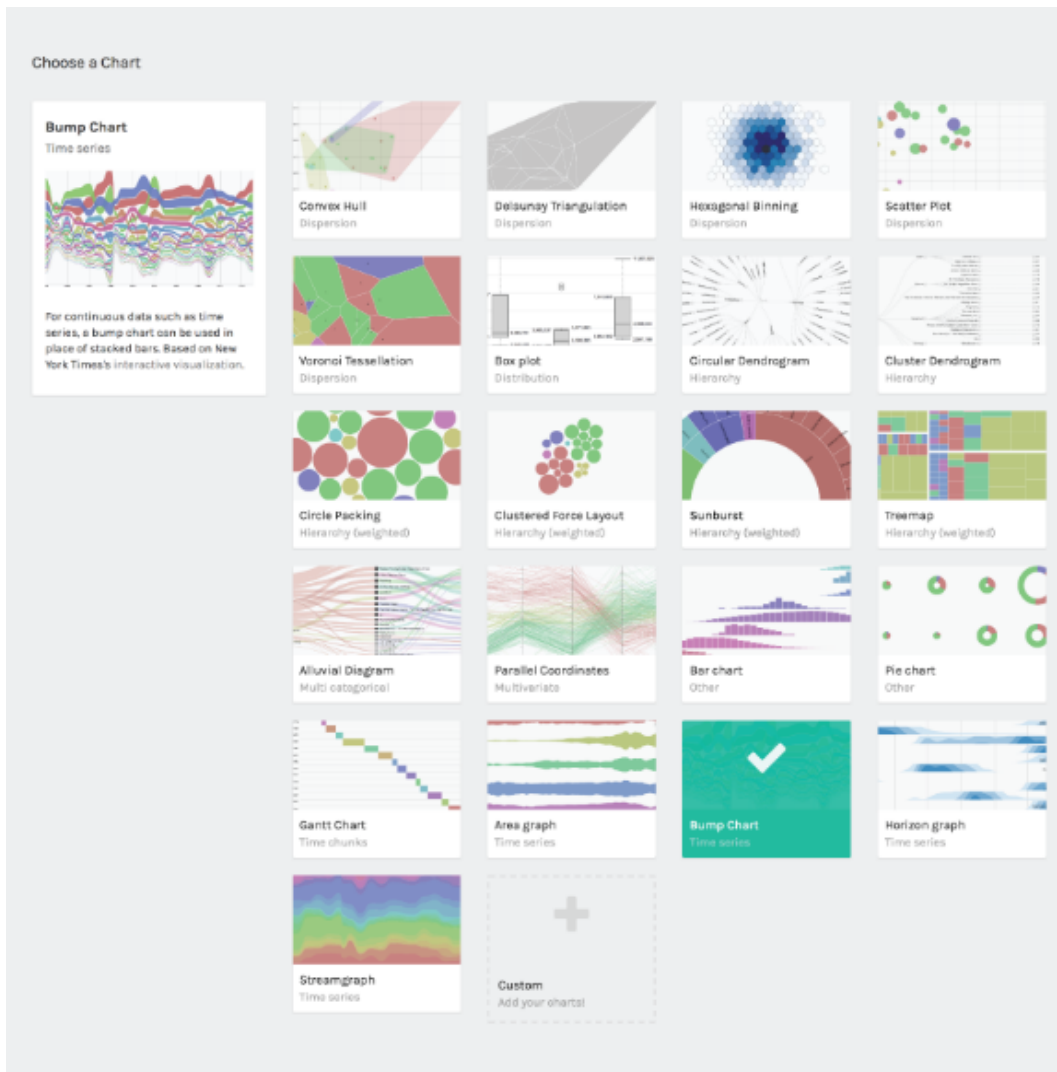
Figur 5: RAWGraphs

5. CONCLUSION

In this paper, we have reviewed some of the available models for data collection, data analysis and data visualisation that are currently adopted by different scientific organisation. The selected tools and platforms are by no means complete but illustrative state-of-the-art approaches to dealing with data impact registration, validation and visualization. In the previous sections, we have provided an analysis of strengths and weaknesses of selected platforms and tools. The models, methodologies and tools analysed, can serve different purposes and applied differently depending on the unit of evaluation, assessment and analysis. Whereas the research registration management platforms serve to capture impact activities, including the impact of activities and outputs from the social sciences and humanities, the available visualisation tools and platforms can present data in different formats for different audiences. In general, a precautionary and responsible application of metrics, tools and visualisations are recommended. The tools and platforms can be used to experiment and represent data from a wide range of sources but they should be considered with care and caution. On rigorous inspection, data sources may not always be robust, which in some instances can be hidden to the external observer if not clearly stated.

References

Adie, E. (2019). Altmetric Blog Gaming altmetrics. Retrieved from
https://www.altmetric.com/blog/gaming-altmetrics/

Boruta, L. (2019). Web-Scale Citation Tracking : Can You Read the Signs. *Medium.Com*, 1–4.
Retrieved from https://medium.com/thunken/web-scale-citation-tracking-can-you-read-
the-signs-98ce0dae5004

Pedersen, D.B., Grønvad, J.F. & Hvidfeldt, R. (2020). "Methods for mapping the impact of social
sciences and humanities—A literature review", *Research Evaluation* 29(1): 4–21 (doi:
10.1093/reseval/rvz033)

Ding, Ying; Mitchell, Stella; Corson-Rikert, Jon; Lowe, Brian; He, Bing (2011). The VIVO Ontology:
Enabling Networking of Scientists (PDF). Retrieved 21 January 2020.

Jacomy, M., Venturini, T., Heymann, S., & Bastian, M. (2014). ForceAtlas2, a continuous graph
layout algorithm for handy network visualization designed for the Gephi software. *PLoS
ONE*. https://doi.org/10.1371/journal.pone.0098679

Mauri, M., Elli, T., Caviglia, G., Uboldi, G., & Azzi, M. (2017). RAWGraphs, 1–5.
https://doi.org/10.1145/3125571.3125585

McEvoy, K. (2015). PlumX Suite Now Available from Plum ™ Analytics ~ Group of Products Are
Designed to Fit the Exact Needs of a Variety of Users ~. Retrieved from
http://www.prweb.com/releases/plumanalytics/plumsuite/prweb12882648.htm

OPERA. (2019). *Dimensions as an OPERA Data Source*.

Spaapen, J., & Van Drooge, L. (2011). Introducing 'productive interactions' in social impact
assessment. *Research evaluation*, 20(3), 211-218.

Thelwall, M. (2018). Dimensions: A competitor to Scopus and the Web of Science? *Journal of
Informetrics*, *12*(2), 430–435. https://doi.org/10.1016/j.joi.2018.03.006

van Eck, N. J., & Waltman, L. (2010). Software survey: VOSviewer, a computer program for
bibliometric mapping. *Scientometrics*, *84*(2), 523–538. https://doi.org/10.1007/s11192-
009-0146-3

Van Eck, N. J., & Waltman, L. (2014). Systematic retrieval of scientific literature based on citation
relations: Introducing the CitNetExplorer tool. *CEUR Workshop Proceedings*, *1143*, 13–20.

Waltman, L., & Van Eck, N. J. (2013). A smart local moving algorithm for large-scale modularity-
based community detection. *European Physical Journal B*, *86*(11).
https://doi.org/10.1140/epjb/e2013-40829-0

ACKNOWLEDGEMENTS