DOI: 10.5281/zenodo.5702265

Triple

Transforming Research through Innovative Practices for Linked Interdisciplinary Exploration

[SEPTEMBER 2021]

Advancing Open Scholarship

D4.3 – GOTRIPLE PROTOTYPE

Version 1 – Draft PUBLIC

H2020-INFRAEOSC-2019
Grant Agreement 863420

The project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 863420

Deliverable Name

Project Acronym:	**TRIPLE**
Project Name:	**Transforming Research through Innovative Practices for Linked Interdisciplinary Exploration**
Grant Agreement No:	**863420**
Start Date:	**1/10/2019**
End Date:	**31/03/2023**
Contributing WP	**WP4**
WP Leader:	**CNRS (Huma-Num)**
Deliverable identifier	**4.3**
Contractual Delivery Date: 02/2020	**Actual Delivery Date: 09/2021**
Nature: Report	**Version 1 :Draft**
Dissemination level	**PUBLIC**

Revision History

Version	Created/Modifier	Comments
0.0	Emilie Blotiere, CNRS (Huma-Num)	First content
0.1	Suzanne Dumouchel CNRS (Huma-Num); Laurent Capelli,CNRS (Huma-Num); Bruno Vandenaert CNRS (Huma-Num)	First review
0.2	Lucas de Santis, Net7; Peter Kraker ,OKMAPS, Simone Kopeink, Know-Center;Gaël Van Weyenbergh, MEOH; Arnaud Gingold, CNRS (OpenEdition)	Second review of technical partners
1	Emilie Blotière, CNRS (Huma-Num); Suzanne Dumouchel, CNRS (Huma-Num), Laurent Capelli, CNRS (Huma-Num), Luca de Santis, Net7	Final version

Table of Contents

Table of Figures

Acronyms

AAI	Authentication- authorization infrastructure
EOSC	European Open Science Cloud
ORCID	Open Researcher and Contributor ID
RI	Research Infrastructure

Publishable Summary

This deliverable is the beta version of the GoTriple prototype. It is an intermediate between the alpha version (March 2021) and the beta+ version that is scheduled for March 2022.

This beta version does not contain all the features and not every innovative service has been integrated yet. It is nevertheless the first release of the public prototype. This beta version is therefore an important intermediate version because it is from this version that the consortium will be able to collect the first comments from users outside in dedicated workshops. It is in line with the common gantt[1] of the project.

The main axes to deliver this prototype have been the following :

- Creating an ETL pipeline
- Deploying the Portal as designed in WP3 which covers the users requirements in a precise co-design process
- Making data available for the innovative services available in this version (3 among the 5 foreseen)
- Enabling resource and data discovery for researchers, people at large
- Integration of the EGI Check-iIn service
- Definition of the API
- Opening the demonstrator

The url to access the GoTriple prototype is : https://gotriple.eu

This document aims to give an overview of the main features available in the beta version and to give the first figures of data imported and the number of aggregators and service providers. Finally, the document sums up what are the next steps to achieve for the beta+ version in March 2022.

[1] https://docs.google.com/spreadsheets/d/1_9FRQD2NjGBOTJKduKTUrlqfPoDLbOLuTEXPbK0dDqQ/edit?usp=sharing

FIGURE 1. PICTURE OF THE FIRST RELEASE OF THE PROTOTYPE

1 THE MAIN AVAILABLE FEATURES

1.1 Publications

5 679 602 documents are available in GoTriple from now on, and the figures evolve continuously.

Among the current number 4 729 397 publications and 119 157 datasets are accessible via a pipeline where data are ingested, classified, enriched and categorised so that they can be easily found and retrieved by the users. The data harvesting system is described in the deliverable 4.1 Technology and Implementation Plan - and Risk assessment, submitted in February 2020[1].

The TRIPLE core pipeline ingest data from aggregators that harvest metadata from providers and repositories. The following 7 aggregators are progressively imported into the platform : DOAB, BASE, DOAJ, EKT, Europeana, FBC Poland, OpenAIRE, Isidore. In addition to the aggregators, the platform also will harvest data from approximately 400 repositories. Both aggregators and providers cover the nine languages of GoTriple (Croatian, English, French, German, Greek, Italian, Spanish, Polish and Portuguese). 27 SSH disciplines are also covered in each of the nine languages.

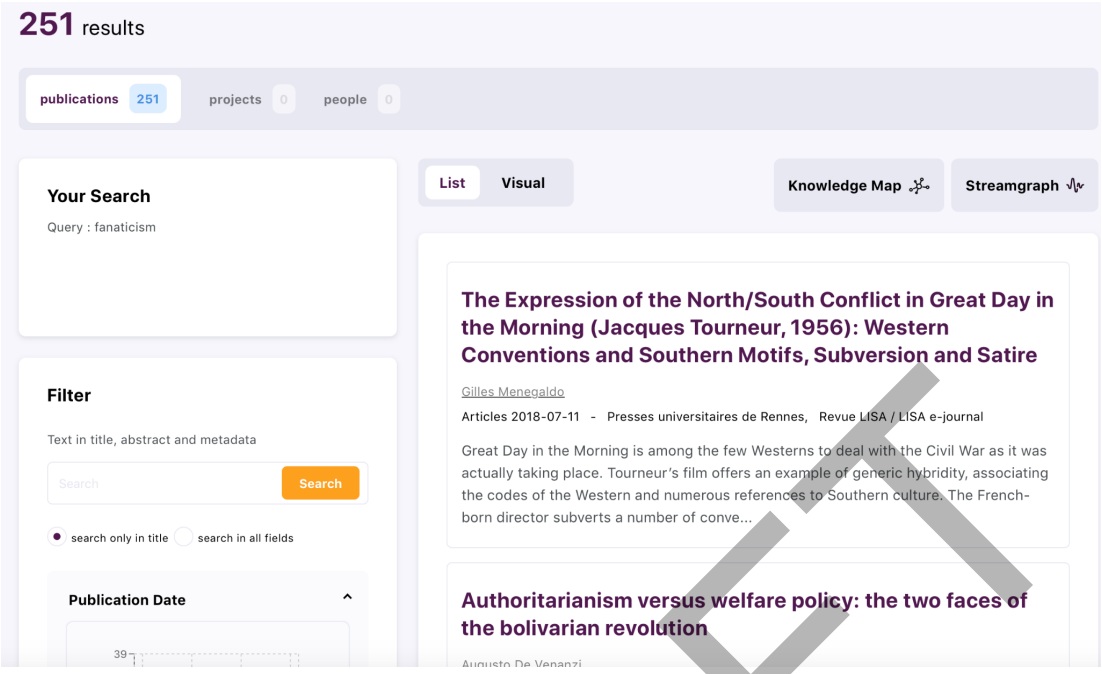

FIGURE 2. SCREENSHOT FROM GOTRIPLE OF THE SEARCH PAGE FOR THE CONCEPT OF FANATICISM

1.2 Multilingual search

GoTriple currently allows multilingual search, which means that users can find publications in a specific language by using keywords in one of the nine languages supported by the platform (Croatian, English, French, German, Greek, Italian, Spanish, Polish, Portuguese). A TRIPLE thesaurus is integrated to the platform containing exactly 2565 concepts available in the nine languages. Documents are annotated with the concepts of the TRIPLE thesaurus in the page of each document. Figure 3 illustrates the result of the concept of fanaticism in Croatian : fanatizam.

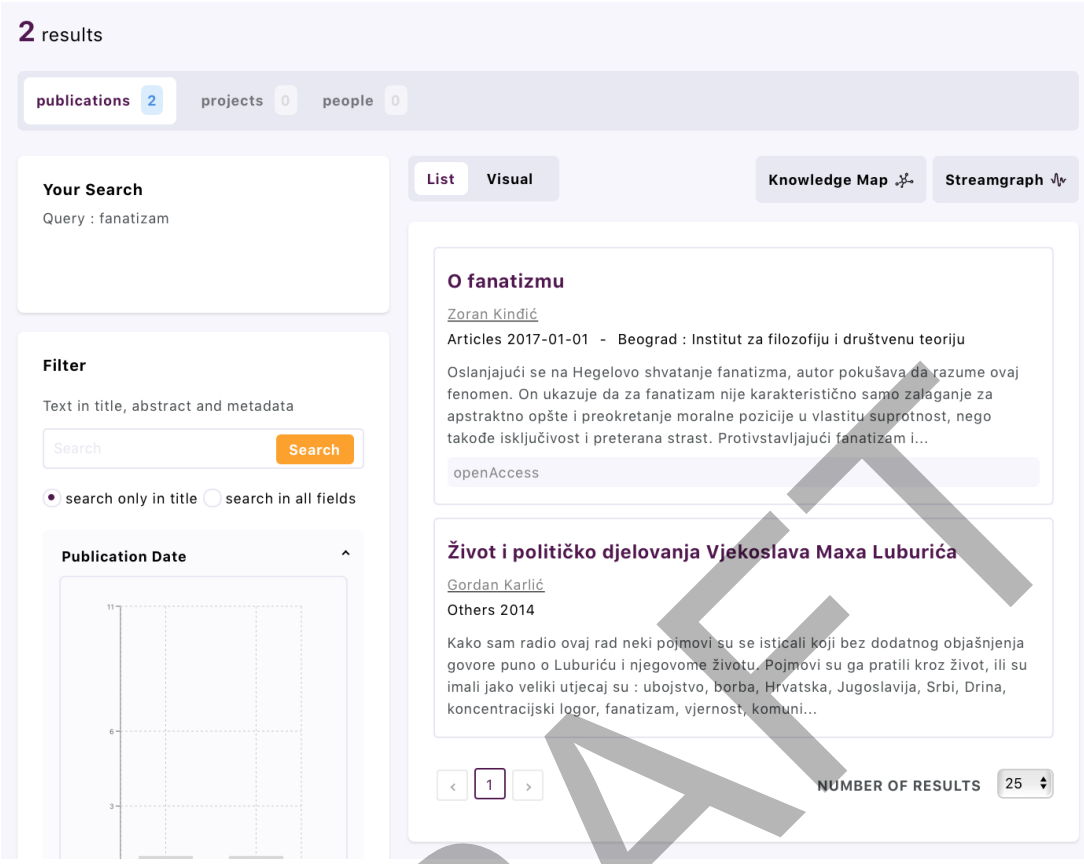

FIGURE 3. SCREENSHOT FROM GOTRIPLE OF THE RESULTS OF THE CONCEPT FANATIZAM IN CROATIAN

1.3 Searching for publications and authors

The search engine allows different filters such as the type of publication, author, year and discipline. Figure 4 shows the results of authors for the concept of fanaticism. The number of publications per author containing this concept is indicated in front of each author. The click on the number of publications per author gives access to the details of the publications (see Figure 5). Figure 6 illustrates the bar chart visualisation of the number of publications per author. By accessing the original page of the document, users can download the available items.

FIGURE 4. SCREENSHOT FROM GOTRIPLE OF THE FILTER AUTHOR FOR THE CONCEPT OF FANATICISM

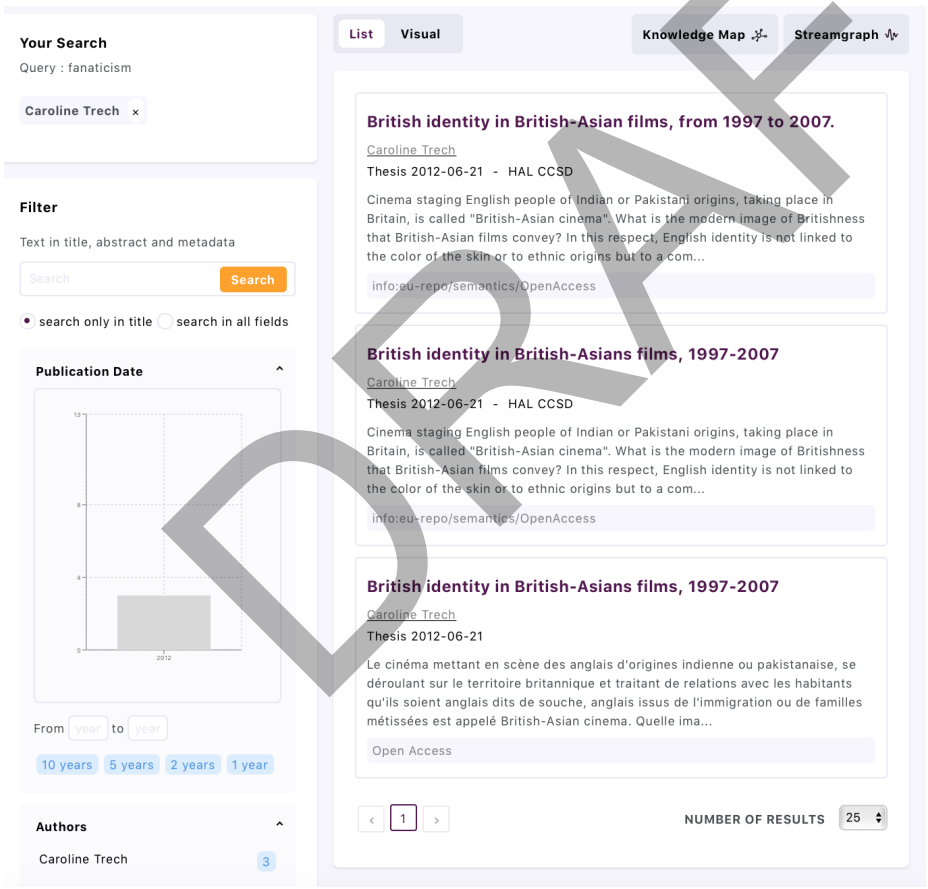

FIGURE 5. SCREENSHOT FROM GOTRIPLE OF THE PUBLICATIONS OF THE AUTHOR CAROLINE TRENCH ABOUT THE CONCEPT FANATICISM

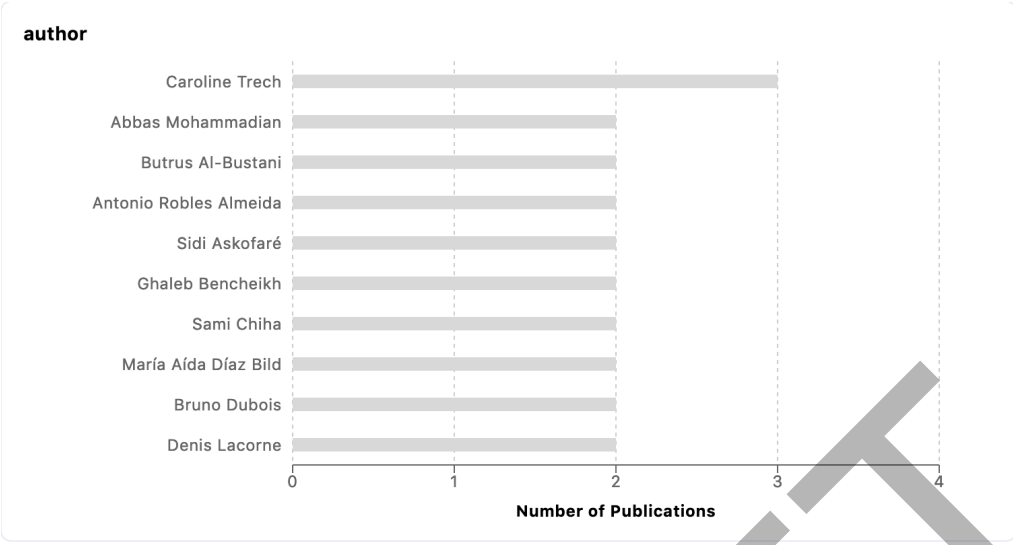

FIGURE 6. SCREENSHOT FROM GOTRIPLE OF THE NUMBER OF PUBLICATIONS BY AUTHOR ABOUT THE NOTION OF FANATICISM

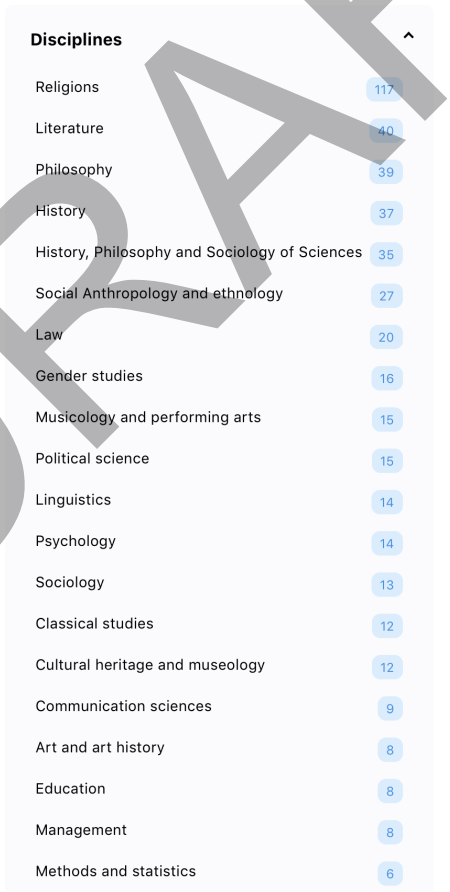

FIGURE 7. SCREENSHOT FROM GOTRIPLE OF THE FILTER DISCIPLINES FOR THE CONCEPT OF FANATICISM

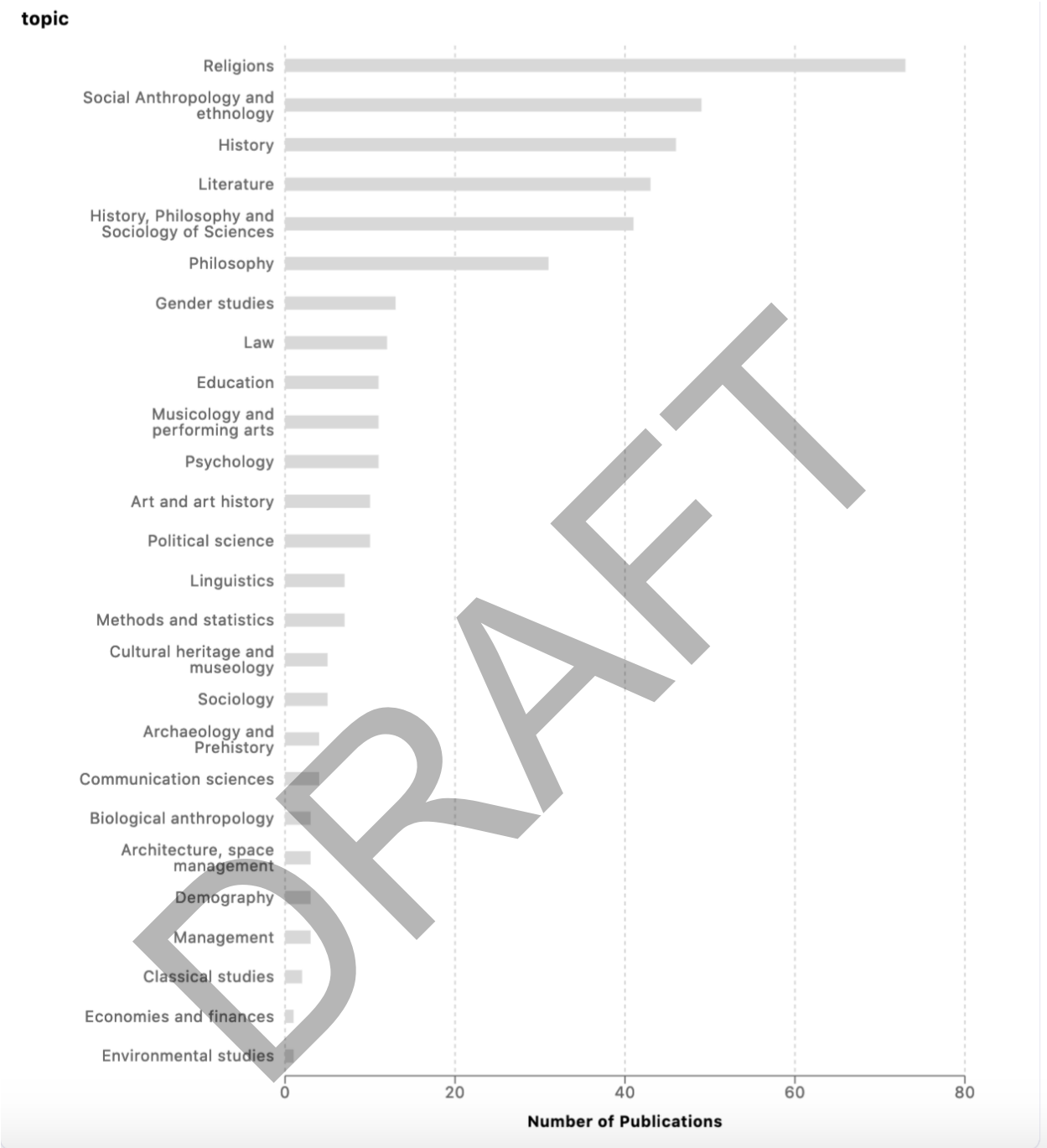

FIGURE 8. SCREENSHOT FROM GOTRIPLE OF THE NUMBER OF PUBLICATIONS PER DISCIPLINE CONTAINING THE CONCEPT OF FANATICISM

1.4 Innovative services

Among the five innovative services represented in Figure 9, three of them are integrated into the beta version of the platform :

- The visualisations
- The discovery system
- The recommender system

They are in the level of tightly integrated services and back-end services.

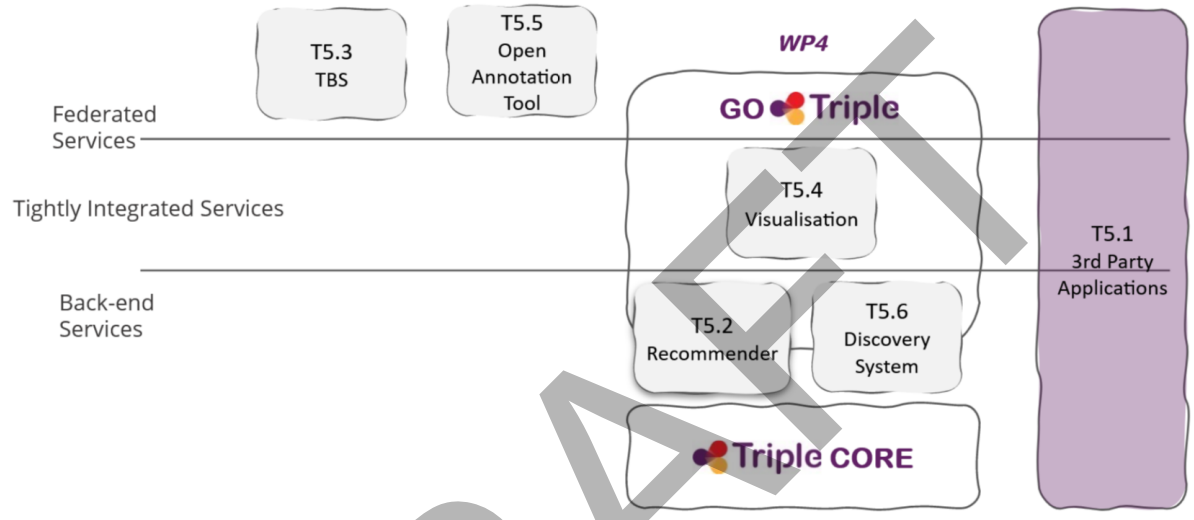

FIGURE 9. INNOVATIVE SERVICES AND THEIR LEVEL OF INTEGRATION (SOURCE D6.2)

1.4.1 Visualisations

Visualisations provide a quicker overview of and insight into a set of research outputs than a pure list-based approach. Two sets of innovative visualisations have been integrated :

- knowledge map and streamgraph.
- diagram components

1.4.1.1 KNOWLEDGE MAP AND STREAMGRAPH

Knowledge map and streamgraph are based on the longstanding open source framework Head Start, which is adapted and extended within TRIPLE for the needs and requirements of the SSH domain. These two visualisations can be embedded in other pages, but they are usually used on their own. The diagram components are not meant to be used as standalone interfaces but to be used on various pages.

1.4.1.1.1. Knowledge map

A knowledge map provides a topical overview of a set of resources by showing the main areas at a glance, and resources related to each area annotated with keywords, comments and tags. This makes it possible to easily identify useful, pertinent information. Research areas are displayed as bubbles. By clicking on one of the bubbles, users can inspect the resources assigned to it (see Figure 11). The size of

the bubbles is relative to the number of resources assigned to it. Closeness of bubbles implies subject similarity. The closer two bubbles, the closer they are subject-wise. Centrality of bubbles implies subject similarity with the rest of the map, not importance. The closer a bubble is to the center, the closer it is subject-wise to all the other bubbles in the map.

FIGURE 11. SCREENSHOT FROM GOTRIPLE OF THE KNOWLEDGE MAP FOR THE CONCEPT OF FANATICISM

1.4.1.1.2 Streamgraph

A streamgraph presents users with a temporal overview of a set of resources over time (see Figure 12 as an example). They are useful for investigating the evolution of keywords over time and to analyse trends in research. The development and features of this tool are described in the deliverable 5.4, Report on the visualisations, submitted in September 2021.

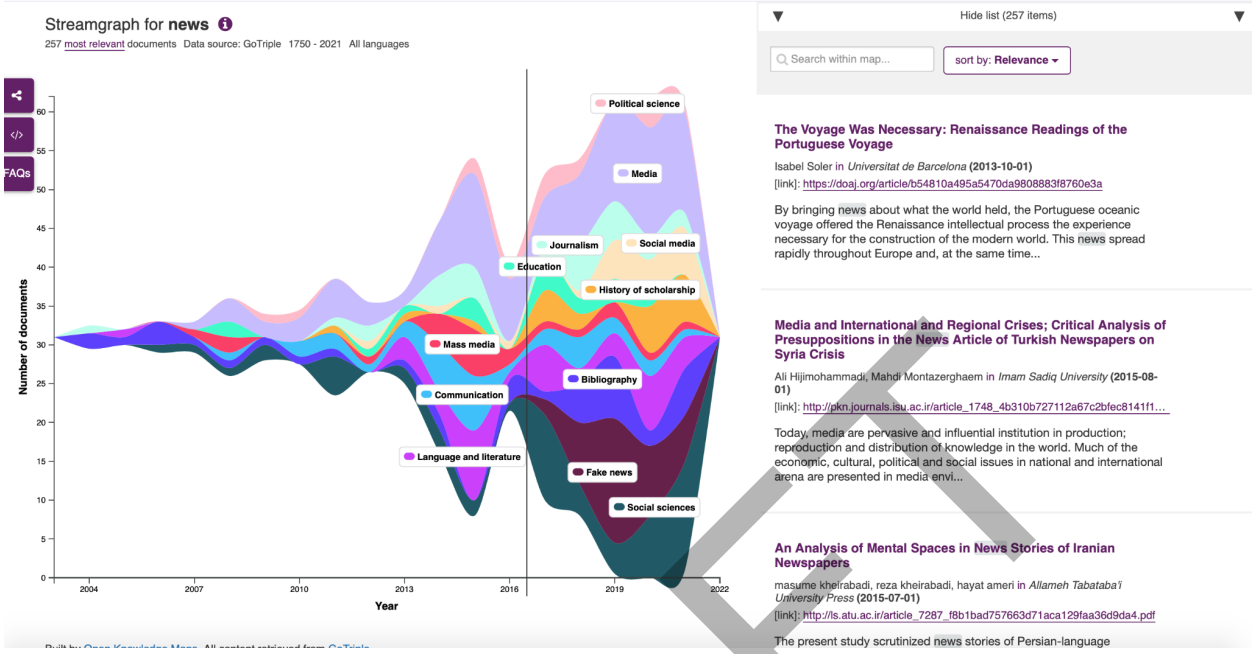

FIGURE 12. SCREENSHOT FROM GOTRIPLE OF A STREAMGRAPH FOR NEWS

1.4.1.2 DIAGRAM COMPONENTS

The diagram components are a set of components based on widely known diagram types. They are designed to be employed in different scenarios and to handle a wide variety of input data. They support limited interaction via popovers and sliders, and they broadcast events that other components can listen to. They are updateable, responsive and mobile-friendly. The diagrams are open source as part of the GoTriple platform repository. Three different diagram types have been implemented: bar chart, line chart, and geo map.

- A bar chart "presents categorical data with rectangular bars with heights or lengths proportional to the values that they represent" (Wikipedia contributors 2021 [8], see Figure 13). The bar chart developed for GoTriple is available both in horizontal and vertical format. In the vertical format, it supports a slider to limit the categories on the x-axis (e.g. the number of years displayed). When a user hovers over a bar, category name and value are displayed.

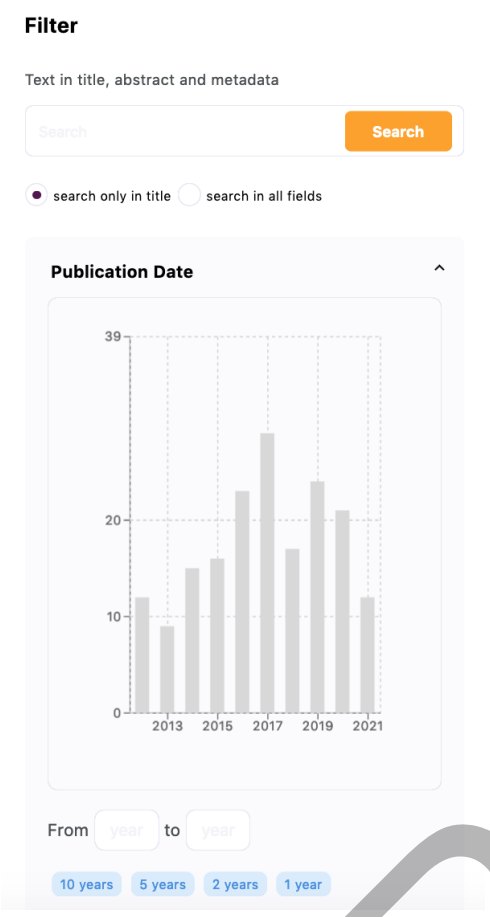

FIGURE 13. SCREENSHOT FROM GOTRIPLE OF THE BAR CHART ABOUT THE PUBLICATION DATE OF THE DOCUMENTS CONTAINING THE CONCEPT OF FANATICISM

- A line chart “displays information as a series of data points called 'markers' connected by straight line segments” (Wikipedia contributors 2021, see Figure 14). The line chart developed for GoTriple is optimized to handle up to seven series of data points in a single diagram. Particular care was given regarding the accessibility of the component for color-blind users: only four easily distinguishable colors are used and further disambiguation is done via line width and dashed lines. A legend is displayed next to the chart to show the series names, but users can also get information by hovering over the markers to see the series name and value at that marker. As in the bar chart, a slider can be used to limit the categories on the x-axis (e.g. the number of years displayed).

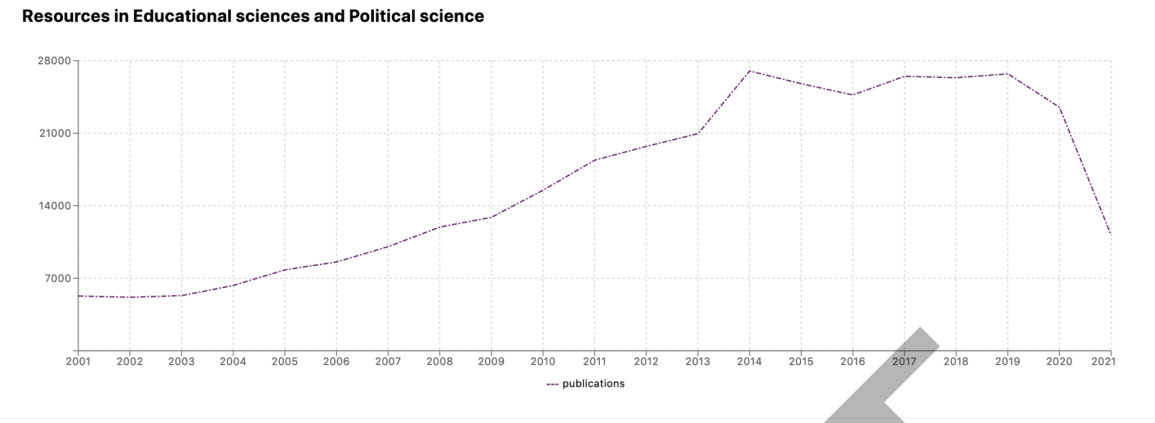

FIGURE 14. SCREENSHOT FROM GOTRIPLE OF A LINE CHART ABOUT THE NUMBER OF PUBLICATIONS PER YEAR OF RESOURCES IN EDUCATIONAL SCIENCES AND POLITICAL SCIENCE

1.4.2 Discovery system - Head Start

The discovery system is based on Head Start, an open source framework that has been developed by Open Knowledge Maps to provide interactive, web-based overviews of research topics, authors, and projects. The overviews are based on related resources, for example, the most relevant publications on a topic, or the articles and datasets published by an author or project[2]. Figure 10 gives an overview of how knowledge maps based on publications are computed.

This technology highlights connections between research outputs using clustering, to help users to sort the relevant from the irrelevant to their research question. This system makes possible the visualisations features described in section 1.4.1 Visualisations and its integration has been the subject of a dedicated deliverable submitted in September 2021 (D.5.6 Discovery system).

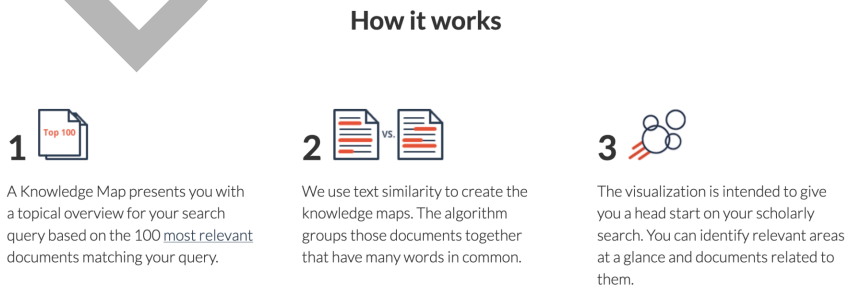

FIGURE 15. A HIGH-LEVEL OVERVIEW OF THE COMPUTATION AND FUNCTIONALITY OF KNOWLEDGE MAPS BASED ON PUBLICATIONS. SCREENSHOT FROM GOTRIPLE OF THE CREATION PAGE OF THE KNOWLEDGE MAP.

[2] Deliverable 5.6 Discovery System - TRIPLE project

1.4.3 Recommender System

The Recommender Systems is a software component that predicts and suggests items of interest to users or user groups. To date, recommendation algorithms mostly use data about users, items and user-item interactions, where users' historical data traces form the main information source of the recommendation logic. The service and its integration into the platform is described in the deliverable 5.2 Recommender System submitted in September 2021.

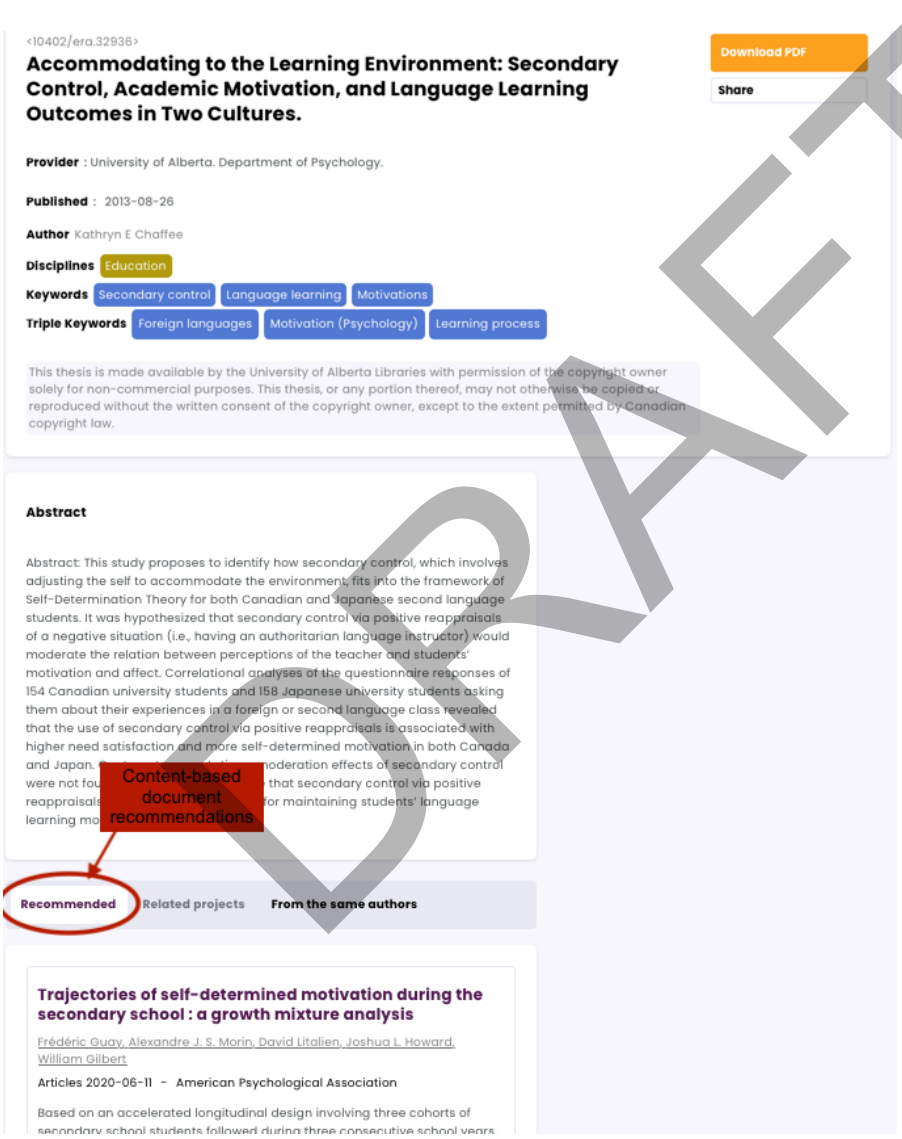

FIGURE 15. SCREENSHOT OF GOTRIPLE SHOWING CONTENT-BASED DOCUMENT RECOMMENDATIONS. THE TAB IS MARKED IN RED (SOURCE D5.2)

2. NEXT STEPS TO ACHIEVE THE BETA+ VERSION

2.1 Additional categories of data

The GoTriple platform will also allow for the discovery of profiles and projects. The data collection process will however, in this case, differ from the one used for publications and datasets. In fact, information on profiles and projects will be, on one hand, automatically retrieved, and on the other hand, created by the users.

2.1.1 Profiles

The profiles represent a complement to the existing database of authors already searchable in the platform beta version. Authors and project's owners constitute a TRIPLE profile with very little information (name, PID). If a user creates an account then she will be able to fill in a lot of other data about herself.

The first addition will come from the automated query of the ORCID[3] database, whenever the ORCID is available in the metadata. This will allow both to enrich the authors' profile and to connect it to publications.

The second addition will come from the users having created their profile through the My GoTriple page. The TRIPLE Data Model allows for a rich description of a person, and the identified users will have therefore the possibility to describe themselves more accurately. In case their ORCID number is not yet known to the system, they will have the possibility to add it in their profile, allowing thus to connect the profile with all the publications recorded in the ORCID database.

The goal of these additions is to increase the level of details of the profiles in order to facilitate the discovery of potential collaborators.

2.1.2 Projects

Like what happens with the profiles, projects will be added to GoTriple through a process either automated or manual.

Looking for projects already funded through broad schemes, GoTriple will first collect data on projects listed in CORDIS[4]. CORDIS stands for Community Research and Development Information Service, and is the European Commission's primary source of results from the projects funded by the EU's framework programmes. The CORDIS data will be cross-checked and completed when necessary with the information available on OpenAIRE repositories. Additional information about the projects will be retrieved from the ORCID profiles, which allow to describe various types of funding. GoTriple will also collect the data about projects funded by

[3] https://orcid.org
[4] https://cordis.europa.eu/about/en

the French national research agency, the ANR[5], and will search for comparable sources within Europe to be added.

TRIPLE also aims at gathering information on projects not yet funded, or in progress. Projects submitted to the Crowdfunding platform will be integrated within GoTriple and made available for discovery. This feature will be developed in March 2022 for the second release of the platform.

Moreover, registered users will be allowed to add projects providing minimal information to avoid duplicates (e.g. based on identifiers). This user-created information will be monitored through controlled lists of funding bodies, similar to the feature currently available on Zenodo.

The consortium is also evaluating the possibility to automatically import data about Citizen Science projects managed through the VERA platform, which is another OPERAS' service. VERA is currently under development in the COESO European research project[6]: the details of this integration will be clarified once the platform is released.

2.2 EGI AAI Check-In service

The EGI Check-in is the User Authentication system (AAI solution) chosen for the OPERAS services, including the GoTriple platform. It supports authentication via multiple different Identity Providers (IdPs) such as the eduGAIN federation. This will also be coupled with the different innovative services linked to GoTriple, allowing users to seamlessly access all the services together with a single sign-on. Within the EGI Check-in service, a Virtual Organisation (VO) for OPERAS has been created and will serve user and group management for all OPERAS services, including GoTriple.

2.3. Innovative services

2.3.1 Recommender System

The recommender system at the time offers only 2 types of document recommendations. At a later stage, when additional data will be available, there will also be peer and project recommendations.

2.3.2 Open annotation tool, Pundit

Once users have been guided to the discovery of the content, the open annotation tool will provide a set of tools for bookmarking, annotating and collaborating on that content. The Pundit web annotator has been extended and customised. A set of detailed APIs will be implemented to allow all information created with Pundit to be reused in GoTriple (and EOSC

[5] https://anr.fr/
[6] https://coeso.hypotheses.org

compliant applications). Pundit has been significantly refactored during this first period of TRIPLE: in particular the integration with EGI Check-in has been implemented, together with several new advanced functionalities, including the support of semantic annotations (a distinguished feature of this tool) and the annotation of PDF documents.
Also the integration with GoTriple has been developed, which enables users to federate their accounts on both platforms: this federation, which will be released shortly, allow them to see Pundit notification from their personal MyGoTriple pages.
Other more sophisticated integration strategies amongst the two services will be identified and implemented for the beta+ GoTriple release of March 2022. The development and process to integrate this service is described in the Deliverable 5.5 Report on the Open annotation Tool and submitted in September 2021.

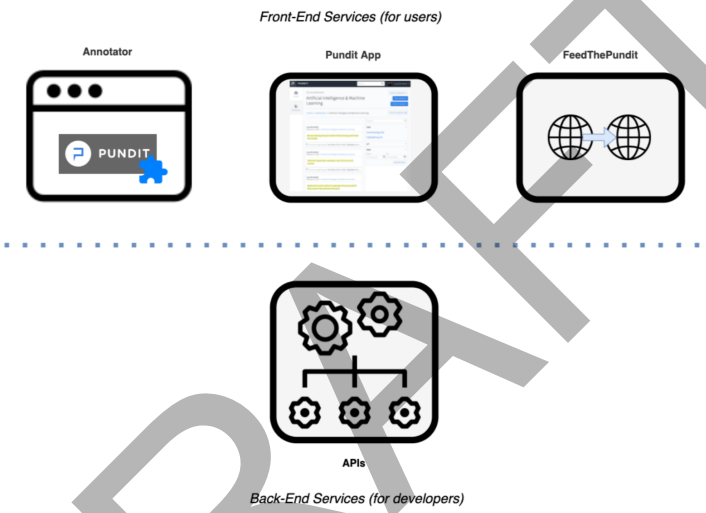

FIGURE 16. PUNDIT'S MAIN SERVICES (SOURCE D5.5)

2.3.3 Trust Building System

The TBS is a referral system meant to help SSH researchers to discover and connect to reliable partners and is one of the "social engines" of the GoTriple platform. The TBS aims to take an approach to social networking where trusted relationships are at the core of the user experience. As a federated service of GoTriple, the service within the GoTriple frontend will be integrated following the three main points :

1. The GoTriple homepage
2. The registered user's notification stream
3. The registered user's profile page

The integration methodology is described in the deliverable 5.3 Report on the Trust Building System Development submitted in September 2021.

2.3.3.1 The GoTriple homepage

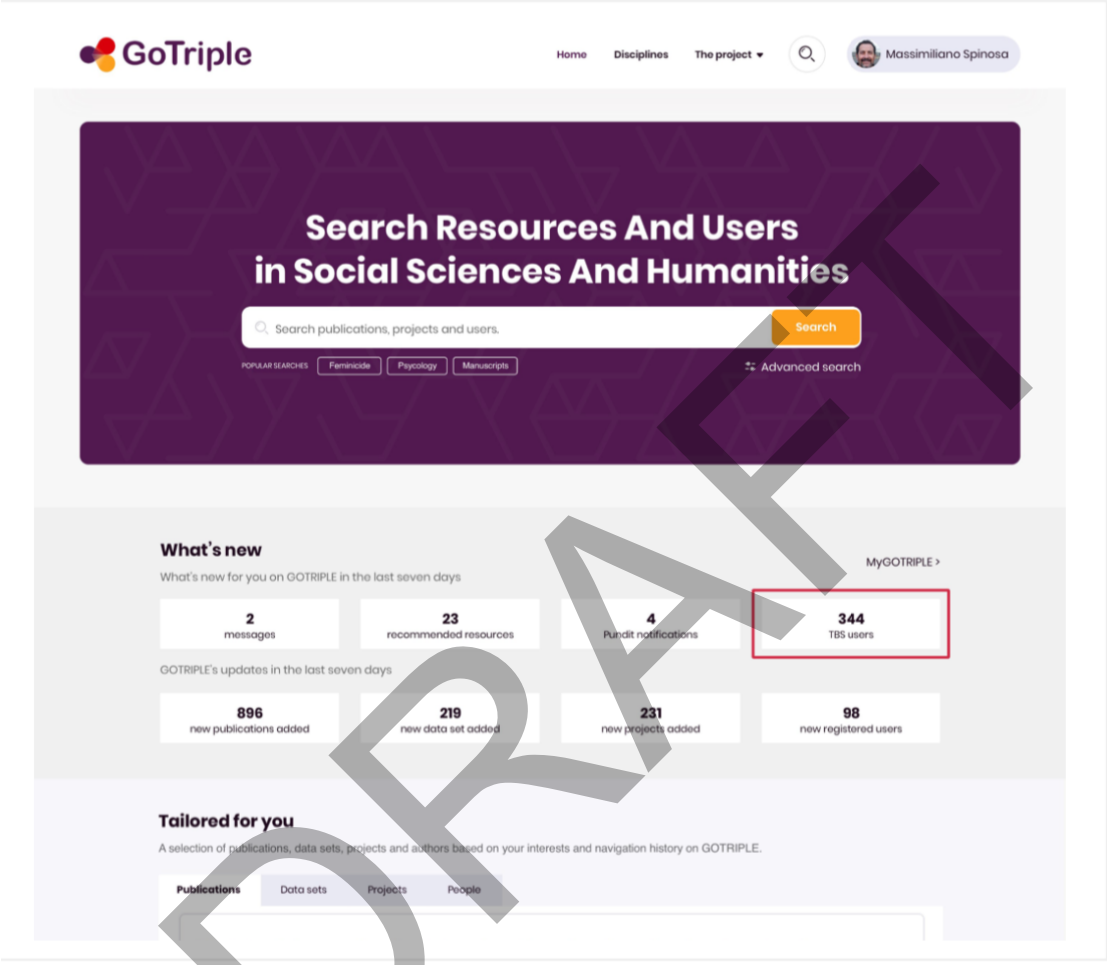

Figure 17. Screenshot of GoTriple homepage showing the main TBS metric (see the red box highlight) (Source D5.3)

2.3.3.2. The registered user's notification stream

A notification stream will be available to the registered users of GoTriple and will display notifications (see figure 19) from various GoTriple services including those of the TBS.

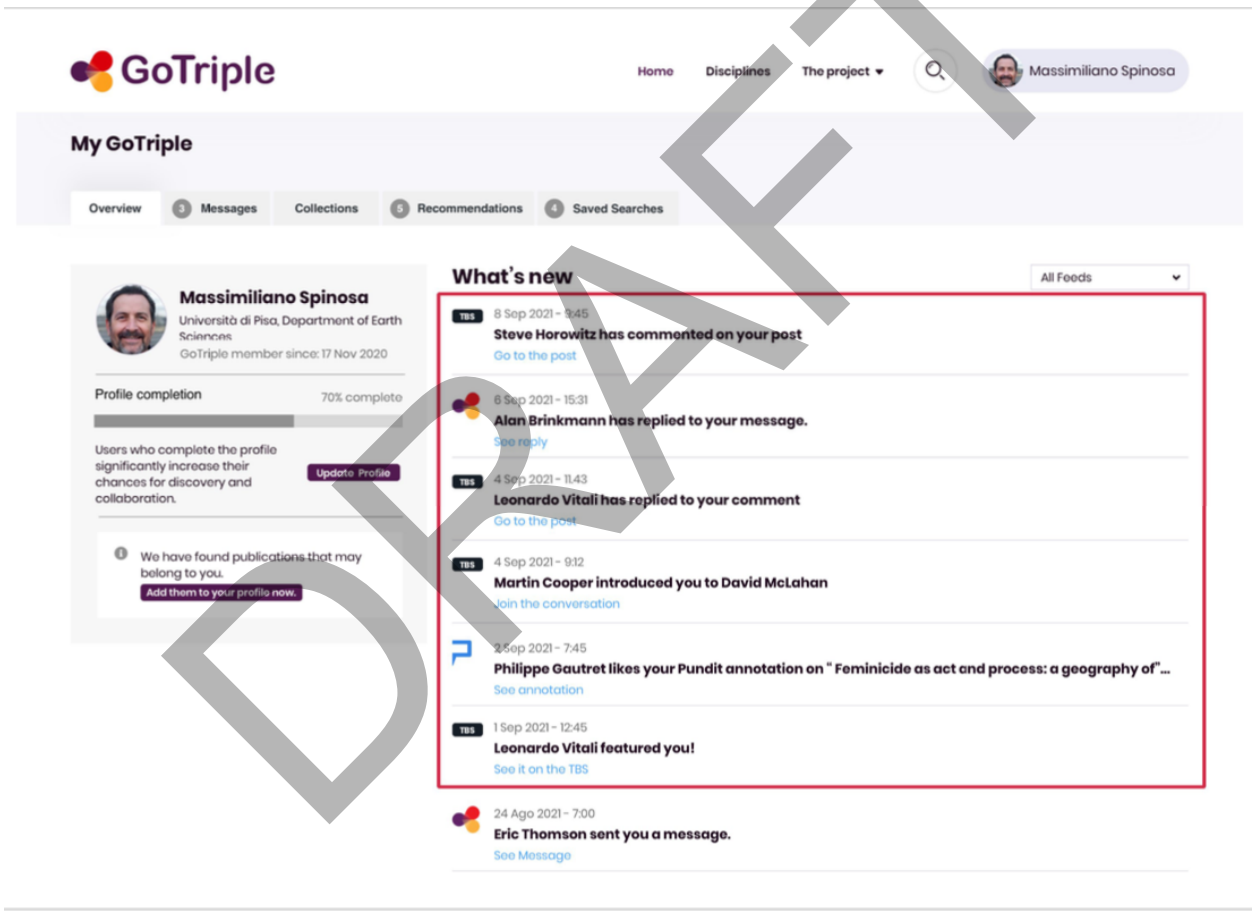

Figure 18. Screenshot of my GoTriple page showing the TBS notifications in the user notifications stream main (see the red box highlight) (Source D5.3)

2.3.3.3. The registered user's notification profile

Finally, a TBS button will be integrated in the registered users' profiles and will provide a direct link to the TBS by opening a new browser tab. The profile settings in the TBS are opened. What still needs to be implemented in the profile settings is a button such as "Link TBS Account to GoTriple" which connects the two accounts. In this way TBS notifications can be communicated to GoTriple and displayed there in the indicated places.

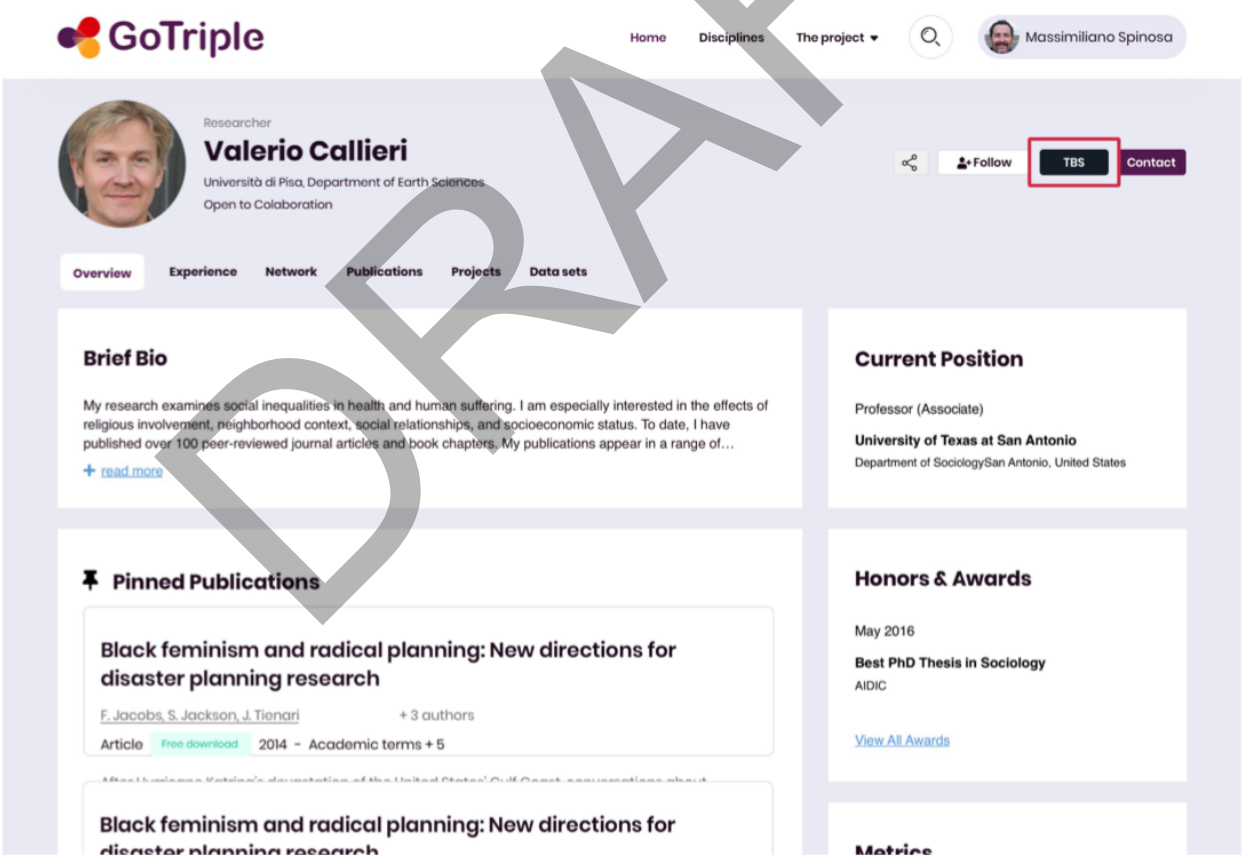

Figure 19. TBS button in GoTriple User ProfilesScreenshot of my GoTriple page showing the TBS notifications in the user notifications stream main (see the red box highlight) (Source D5.3)

2.3.4. CROWDFUNDING PLATFORM

The crowdfunding platform is developed with the selected provider Wemakeit[7]. A timeline has been defined, which will lead to the official launch of the service in January 2022. The work to integrate this tool to the platform is described in the deliverable 5.1-Report on Third-Party Applications integration-1.0, submitted in September 2021. The purpose of this service will be to help to fund specific projects led by researchers in Social Sciences and Humanities. In this perspective, it is expected to strengthen the relations between the researchers and the citizens and to develop scientific projects that are particularly relevant or interesting for citizens. This crowdfunding service is seen as an exploratory tool to increase the societal impact of SSH science.

REFERENCES

[1] Deliverable 4.1 Technology and Implementation Plan - and Risk assessment
[2] Deliverable 5.4 Report on the visualisations
[3] Deliverable 5.6 Discovery System
[4] Deliverable 5.2 Recommender System
[5] Deliverable 5.5 Report on the Open annotation Tool
[6] Deliverable 5.3 Report on the Trust Building System Development
[7] Deliverable 5.1Report on Third-Party Applications integration-1.0
[8] Wikipedia contributors. (2021, June 23). Bar chart. In Wikipedia, The Free Encyclopedia. Retrieved 16:56, September 29, 2021, from https://en.wikipedia.org/w/index.php?title=Bar_chart&oldid=1029967372

[7] https://wemakeit.com