

Powering Scientific Research with Collaborative Agility

Scientific research is knowledge-dependent and collaborative in nature. Timely availability of relevant information and global collaboration among researchers is a necessity for research in today's global world. The need for intelligent and readily accessible information by the scientific community has forced scientific information providers to leverage technology-led advancements in electronic research. Powering online channels with accurate and relevant scientific information will allow scientific researchers to reduce time spent in knowledge acquisition and devote more time to their core research, which in turn will help in creating scientific breakthroughs.

This paper presents the evolution of scientific research in keeping with the use of electronic media, the challenges faced by scientific information providers and TCS's framework to achieve Collaborative Agility for the scientific information providers. This will enable the scientific information providers monetize information; help researchers collaborate and deliver an effective knowledge platform for the scientific community.

About the Author

Debasish Datta

Debasish Datta is a consultant for TCS Media & Information Services and leads the Information Services CoE. He has experience of working with several Information Services companies and has helped them in their transformation journey in achieving Content Agility. He can be reached at debasish.datta@tcs.com.

Table of Contents

1. Introduction	4
2. Changes in Scientific Research with Internet	4
3. Challenges with Increasing Online Reach	6
4. Collaborative Agility	8
5. TCS Solution for Collaborative Agility	9
6. Conclusion	10
7. Keywords	10

Introduction

Scientific research is the backbone for inventions that have over the years contributed to the betterment of life for all of us. It is inherently knowledge driven, and knowledge is accumulated from published research and the expert interpretations of their own research. By nature, it is highly collaborative and the community leverages peer expertise to enhance their knowledge. Collaboration among scientific researchers increases productivity and Big Science often requires collaboration (De Solla Price, 1963).

Scientists consume an enormous amount of information as a part of their day-to-day research activities. Studies reveal that physicists read an average of 204 articles every year, while chemists read approximately 276 articles each year (Tenopir & King, 2001). Scientific researchers and medical workers read three times as many journal articles as legal, management, and sales professionals (Belefant-Miller & King, 2001).

Deep penetration of Internet globally has changed the landscape of scientific research from being primarily dependent on print journals to electronic research. Scientific information providers and scientific publishers have created digital products to address their consumer demands. Huge content available only in print in libraries worldwide, stored in legacy systems in non-granular structures and rapidly growing new sources of scientific content online pose strong challenges for the scientific information providers to ensure comprehensiveness, accuracy and relevance of scientific information. Scientific information providers should leverage efficient content management and online collaboration platforms to make online channels more useful for scientific research. This will drive adaptation of online channels, make information sharing easier and faster among the scientific community and create powerful platforms for collaboration.

Changes in Scientific Research with Internet

During the pre-Internet days, printed scientific journals and books were the primary source of scientific information. Seminars, conferences and personal networking were the only means of collaboration among the scientific community. Libraries were the primary keepers of scientific knowledge in books and journals. Librarians, well qualified in the art of book management, would ensure that books, journals are stacked up in the correct shelf and easily accessible. Extensive peer reviews of the articles/thesis were done for correctness and originality to ensure scientific knowledge is not plagiarised. The cycle time from research to publishing was long, resulting in ideas being made available long after they were actually conceived. Lot of valuable time was spent by researchers in information acquisition, thus limiting the time available to do core research.

ARPANET, the predecessor of Internet was born in University of California (UCLA) and the scientific and academic communities were the first users of this significant technological invention. The deep penetration of Internet made information sharing and availability much easier among the global scientific community. Scientific publishers started exploiting online channels to provide information. This shift from print to online resulted in major changes in scientific knowledge management and discovery. As the demand for speedier information dissemination grew, the online channels became the primary means of scientific information. More and more devices were hooked on to the Internet and the information was required to be made available across all delivery channels.

Even with the benefits of Web 1.0, the information flow online was mostly one sided. Web 2.0 changed all that. The scientific community, relying heavily on effective collaboration has started to adopt Web 2.0 platforms. However, the real value of the technology advancement will be to marry the scientific content repository with the Web 2.0 collaborative platforms. The scientific information providers want to ensure that they are not left behind and have started transformational initiatives to monetise their content better and extend the online product-set across multiple delivery channels seamlessly.

	Pre-Online	Online - Web 1.0 (Additional to pre-online mediums)	Online -Web 2.0/3.0 (Additional to Web 1.0 mediums)
Information Medium	Printed Books, Journals	Online journals, newsletters, alerts, mails, websites/portals	Blogs, wikis, bulletin boards, discussion forums
Primary Research Place	Library	Anywhere with Internet connection	Anywhere with Internet connection
Information Access Mode	Printed journals or books. Use library access methods and Bibliography to locate information	Research online. Full text, keyword search, bibliography citations	Personalised information, ontological navigation
Research Funding Information	Mainly available through personal networking and print notifications	In addition to personal networking, information is available in websites/ portals, electronic notifications.	In addition to Web 1.0, available through online community networks, personalised alerts.
Community Networking	Personally attend conferences, seminars, meetings and events	In addition to personal networking, collaboration using email.	Online Community Platforms accessible through all channels
Peer Review	Peer review in the community network	Extend online through editorial reviews, expert comments on website	Web 2.0 platform - blogs, article reviews, discussion forums
News	Newspapers, Journal news	Scientific websites, news aggregators, emails	Community forums, mashups, personalised news, community news

Table 1: Evolution of research mediums with Technology

Challenges with Increasing Online Reach

Availability of scientific information electronically has a positive impact on the scientific community as information can be obtained quicker and collaborated beyond geographic boundaries. Thereby, they can spend more time in their core research and improve their productivity and probability of technology breakthrough. However, these positive effects and the high dependence on electronic mediums have only increased the challenges for the scientific information services providers. Some of the key challenges are described below:

Huge Print Assets

With the increasing use of online channels by scientific researchers, the huge volume of scientific documentation only available in print in libraries are being neglected. The information contained there is valuable for the researchers and needs to be digitised to ensure that scientific knowledge is not lost. Digitisation and content tagging of millions of pages is associated with significant costs. A lot of libraries and publishers have started digitising, but there are challenges in funding such big conversions globally. There are also legal challenges for content ownership and royalty. Technology needs to make digitisation and conversion to standard formats cheaper.

New Media Content

Higher Internet bandwidth availability has increased the usage of rich media content like pictures, audio and video across the web. Scientific Information can be better represented using rich media. However, processing the huge datasets in scientific documents and extracting information from the unstructured content is a significant challenge.

Information Overload

Increased use of the Internet is creating more and more scientific information contributors across the globe. The number of online publishers is significantly increasing every year. The user-generated content continuously increases the volume of information available. This is causing an information overload for the researcher, as there is more material to go through without the assurance of its authority and originality. Technology needs to enable processing of the huge datasets, link with the existing information and render them in a way that makes it more usable to the scientific community.

Information Accuracy

With an increase in the number of online contributors and user-generated content, everyone has a chance to contribute, which at times can lead to inaccurate information and plagiarism. With shortening review cycles due to demand for quick publishing, the scientific community is apprehensive that the peer review process may be sometimes compromised. Technology needs to ensure that advanced tools are available with publishers and researchers to quickly eliminate/exclude inaccurate content.

Information Access

The content stored in legacy content platforms are mostly in monolithic structure and do not render themselves to be easily searched. With the volume of data increasing significantly, the access mechanism needs to be efficient enough to deliver the results fast. Technology needs to ensure that advanced data mining tools with ontology information are available to drill down the information. There is a need for semantic technologies so that the researchers can locate the contextual information easily.

Multi-channel Delivery

Today, researchers are tech-savvy and extensively use digital devices. They prefer specific devices to deliver information. Information Service providers now are required to deliver in all forms of delivery channels including mobile devices, PDAs, eReaders and RSS feeds at the same time, so that they can be more productive. The content stored must be granular and easily capable of being rendered through multiple delivery channels. Technology needs to ensure that repurposing of content can be done quickly across multiple devices.

Increased User Expectations

With abundant information available, the scientific community expects to get more than just the data facts available in print. They want actionable intelligence to be made available fast. Search results should show answers and not just links. They want to realise the value to be created by online platforms and widen the network of researchers, create more collaboration among the community and produce more authoritative content (user-generated content) quickly. Technology needs to discover the hidden value of data and provide accurate intelligence timely. Collaborative platforms should allow better collaboration and help save precious time that can be put to use in making scientific breakthroughs.

Collaborative Agility

Collaborative Agility is the solution to address the content challenges, build collective intelligence and effective community collaboration for the scientific community. Broadly, it can be defined as a framework that allows:

- content to be loaded from any source and any format (including digitised print archives)
- content to be stored maintaining the inherent scientific document structure (as XML)
- content to be transformed, accessed and analysed quickly
- content to be delivered in multiple formats/channels
- platform for online community collaboration, use and for creating content online
- creation of infrastructure to monetise content

The conceptual framework of Collaborative Agility is provided here. It addresses the challenges faced by scientific information providers.

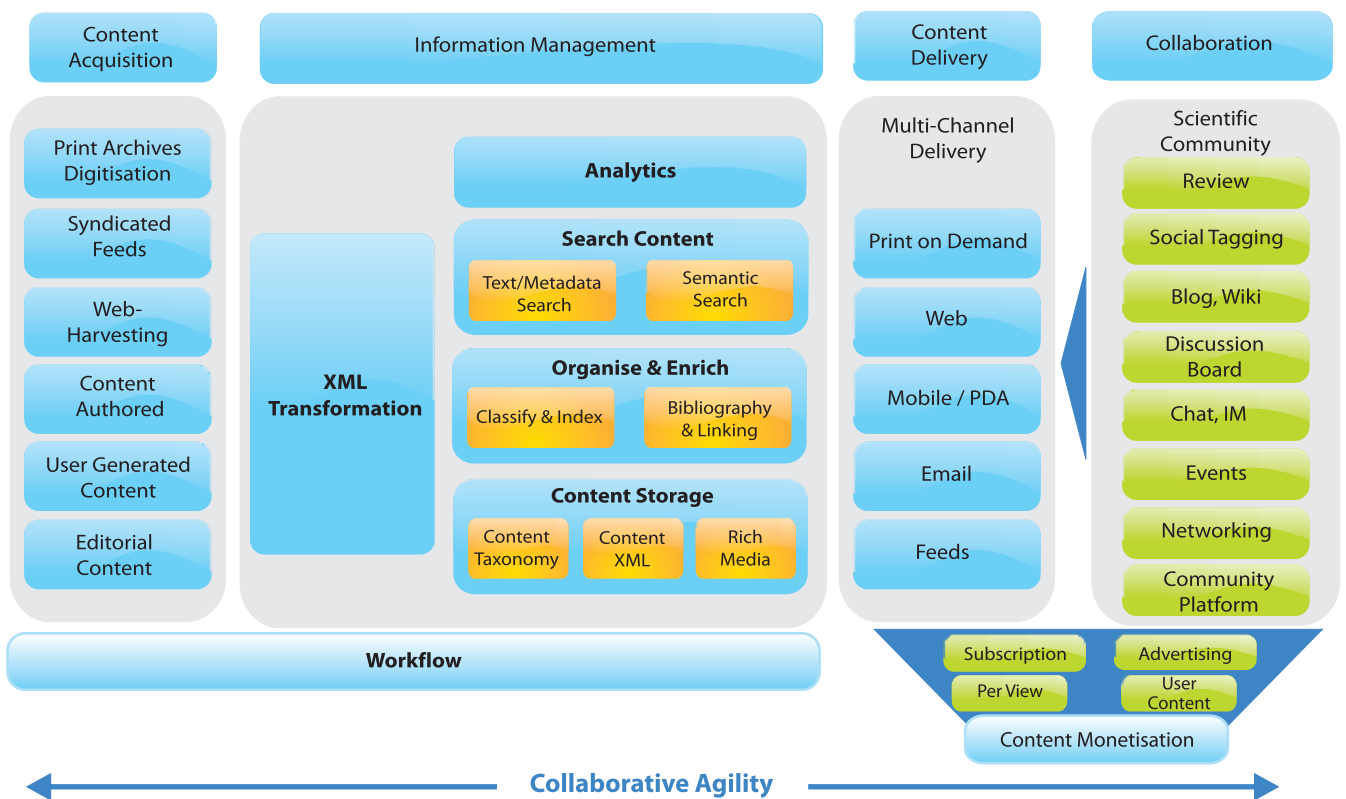


Figure 1: Collaborative Agility Framework

TCS Solution for Collaborative Agility

TCS understands the challenges faced by the scientific information providers. It has combined its technology expertise in Content Agility and Web 2.0 with its experience in providing solutions for scientific information services to create a solution offerings framework for Collaborative Agility. The solution framework uses the products from its alliances/partners like Mark Logic, TCS content products and service offerings. The framework uses open standards and also leverages open source products with the goal to make it pluggable to the existing infrastructure. The framework contains components that TCS delivers as services to make the overall solution cost-optimised and scalable.

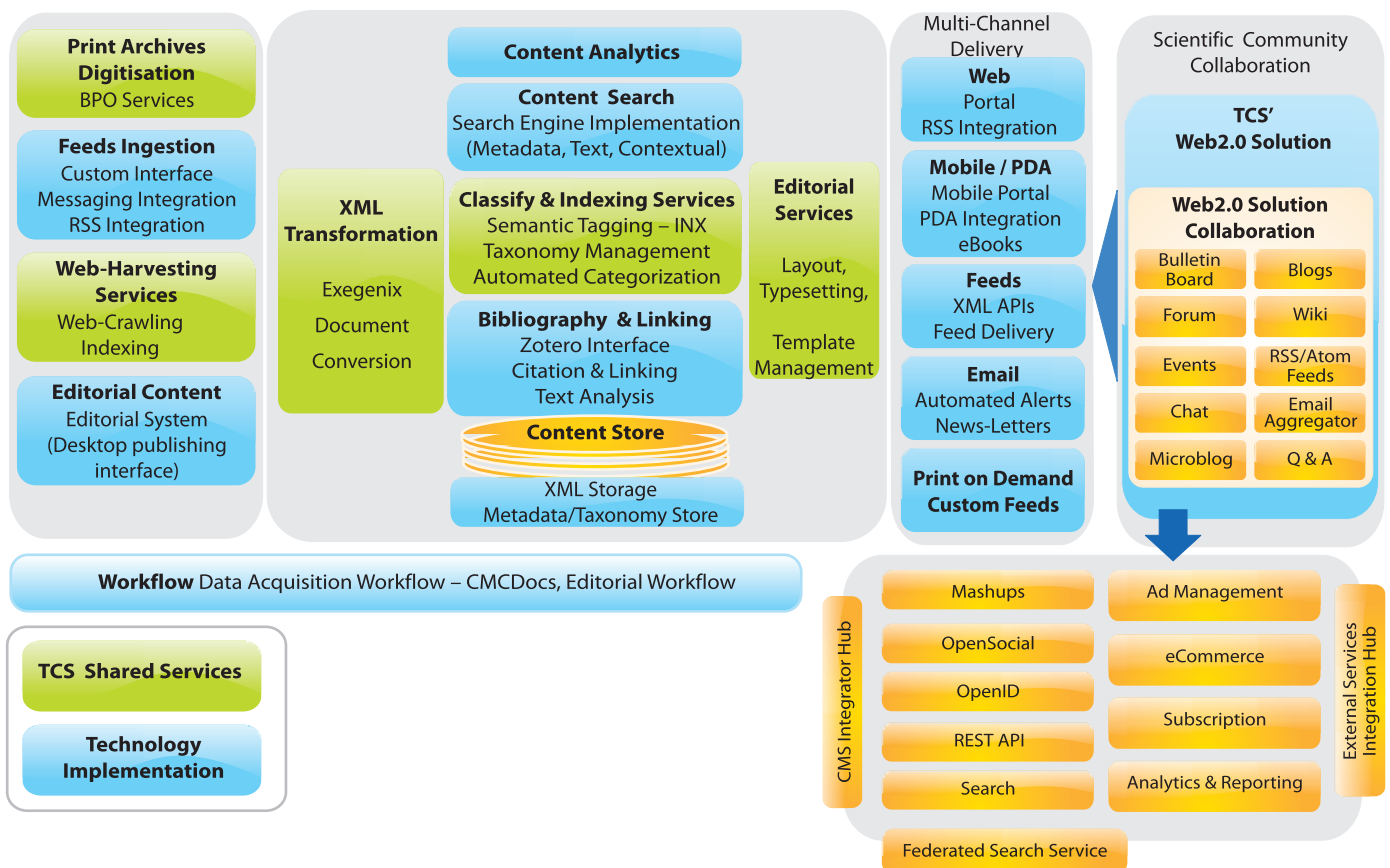


Figure 2: TCS Solution Framework for Collaborative Agility

- Exegenix – Automated Content conversion from documents in pdf/ word /html to XML.
- CMCDocs – Data Acquisition Workflow with inbuilt modules for OCR/Scanning interfaces.
- INX Semantic Tagging – Framework and components developed by TCS Innovation Labs for automated semantic linking, categorisation based on ontology.
- Web 2.0 Collaboration Framework – TCS Open Source Collaboration Framework with Web 2.0 tools and assets for scientific online collaboration.

The salient capabilities of the framework:

- 1) **Digitisation and XML Conversion:** Scientific artifacts will be digitised and transformed to XML cost-effectively. This will allow preservation of important articles and also increase the content richness, and help monetise these legacy assets.

- 2) **Web-Harvesting:** Web-content will be captured intelligently and then transformed into XML so that information can be extracted and linked with the existing data sets.
- 3) **Native Content Structure (as XML):** Scientific documents are unstructured. They are mostly text but also contain large number of pictures/diagrams and even video/audio. The positions of the pictures and the contextual meaning of the words in the sentence are essential to understand the correct meaning. This framework will store the document in the XML structure and allow the native hierarchical content to be maintained. The pictures and the video will be tagged and their context within the document captured.
- 4) **XML Content Repository:** XML content repository stores the content in the native structure and allows content to be retrieved using XQuery. Native XML interface allows content to be retrieved faster.
- 5) **Automated Categorisation and Linking:** Categorise and link content with semantic technologies. Automated ontological classification and categorisation will un-cover the co-relations which otherwise would be difficult to find out manually. There is a provision to manually review the categorisation using the integrated workflow and exceptions corrected. The semantic technology will self-learn and be able to identify such patterns after inclusion. Follow open standards for bibliography linking.
- 6) **Contextual Search:** Leverage the XML content repository and semantic search to provide contextual response to questions. It has the ability to do metadata search, full text search and rich media search. Rich media search is based on XML transcripts and metadata tagged. UI framework will allow the ontology hierarchy and linking to drill down or navigate to related information.
- 7) **Multi-channel Delivery:** Content can be easily repurposed to create tailored content for multiple delivery channels.
- 8) **Community Platform:** Uses the open Web 2.0 platform to build a collaborative framework for the scientific community. The community can leverage the platform to do personalised information dissemination, peer-review, host online webinars, networking, chat, blogging/bulletin boards for richer online collaboration. Social tagging and collective knowledge can be harnessed to provide intelligent contextual results.

Conclusion

Technology can help integrate scientific communities globally and make information available at the right time and right place. Easy accessibility of accurate information and online collaboration platforms will allow researchers to easily access the information they need. Scientific information providers need to make their content agile and provide mechanisms to make the relevant content quickly accessible. They need to provide online collaboration platforms for the scientific community and integrate their online database with collaboration platforms to make it more useful for researchers. TCS Collaborative Agility provides a platform to introduce scientific content agility and a platform to monetise the content.

Keywords

Collaborative Agility, Content Agility, Collaboration, Web 2.0, information services, scientific research



About Tata Consultancy Services (TCS)

Tata Consultancy Services is an IT services, business solutions and outsourcing organization that delivers real results to global businesses, ensuring a level of certainty no other firm can match. TCS offers a consulting-led, integrated portfolio of IT and IT-enabled services delivered through its unique Global Network Delivery Model™, recognized as the benchmark of excellence in software development.

A part of the Tata Group, India's largest industrial conglomerate, TCS has over 143,000 of the world's best trained IT consultants in 42 countries. The company generated consolidated revenues of US \$6 billion for fiscal year ended 31 March 2009 and is listed on the National Stock Exchange and Bombay Stock Exchange in India.

For more information, visit us at www.tcs.com

Contact us at
global.me@tcs.com

All content / information present here is the exclusive property of Tata Consultancy Services Limited (TCS). The content / information contained here is correct at the time of publishing. No material from here may be copied, modified, reproduced, republished, uploaded, transmitted, posted or distributed in any form without prior written permission from TCS. Unauthorized use of the content / information appearing here may violate copyright, trademark and other applicable laws, and could result in criminal or civil penalties.

Subscribe to TCS White Papers

TCS.com RSS: http://www.tcs.com/rss_feeds/Pages/feed.aspx?f=w

Feedburner: <http://feeds2.feedburner.com/tcswhitepapers>

TATA CONSULTANCY SERVICES

www.tcs.com