# The dynamics of beamline configurations and user communities

## Quantitative studies of Big Science publications

Söderström, Kristofer Rolf

2023

*Document Version:*
Publisher's PDF, also known as Version of record

Link to publication

*Citation for published version (APA):*
Söderström, K. R. (2023). *The dynamics of beamline configurations and user communities: Quantitative studies of Big Science publications*. [Doctoral Thesis (compilation), Department of Arts and Cultural Sciences]. Department of Arts and Cultural Sciences, Lund University.

*Total number of authors:*
1



LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00

*Kristofer Rolf Söderström*

# The dynamics of beamline configurations and user communities

## QUANTITATIVE STUDIES OF BIG SCIENCE PUBLICATIONS

LUND STUDIES IN ARTS AND CULTURAL SCIENCES 31

# THE DYNAMICS OF BEAMLINE CONFIGURATIONS AND USER COMMUNITIES

# The dynamics of beamline configurations and user communities

## Quantitative studies of Big Science publications

Kristofer Rolf Söderström

Lund University

LUND STUDIES IN ARTS AND CULTURAL SCIENCES 31

Lund Studies in Arts and Cultural Sciences is a series of monographs and edited volumes of high scholarly quality in subjects related to the Department of Arts and Cultural Sciences at Lund University. An editorial board decides on issues concerning publication. All texts have been peer reviewed prior to publication.

Lund Studies in Arts and Cultural Sciences can be ordered via Lund University: https://www.ht.lu.se/en/series/lsacs/

Faculty of Humanities and Theology
Department of Arts and Cultural Science

ISBN: 978-91-987325-4-2 (print)
978-91-987325-5-9 (pdf)
Lund Studies in Arts and Cultural Sciences 31
ISSN: 2001–7529 (print) 2001–7510 (online)

Coverdesign by Johan Laserna
Typesetting by Jonas Palm, Media-Tryck, Lund University
Printed in Sweden by Media-Tryck, Lund University, Lund 2023

Nordic Swan Ecolabel, 3041 0903

# Table of Contents


Acknowledgements 7

Abstract 11

List of Articles 13

Author's contribution to the articles 15

Abbreviations 17

1. Introduction 19
   - Aim, objectives, and research questions 22
   - Sections of the thesis 24
   - Placing the thesis: Information and Big Science 27
   - The facility and its beamlines 31

2. Literature review 37
   - User-oriented Big Science 37
   - Quantitative studies of Big Science publications 42

3. Theory 45
   - The beamline 46
   - Beamline space 48
   - Beamline infrastructure 50

4. Method 55
   - Data 55
   - Empirical approach 57

5. Summary of articles 61
Article I. Generic instruments in a synchrotron radiation facility 61
Article II. The structure and dynamics of instrument collaboration networks 63
Article III. Global reach, regional strength: Spatial patterns of a big science facility 65
Article IV. Who did what? Creating structured data from acknowledgement texts with large language models 66

6. General discussion 69
Beamline spaces 69
Technological configurations 71
Infrastructural adaptation and visibility 74
On metadata 76
Limitations 78

7. Concluding remarks 81

References 83

Article I 95

Article II 121

Article III 143

Article IV 163

Lund Studies in Arts and Cultural Sciences 185

# Acknowledgements

The research behind this article was funded by the Swedish Research Council, grant no 2018-01091. It is part of the project ‘Small science on big machines: Patterns of collaboration and use of contemporary large neutrons and x-rays research infrastructures.’

It is quite a task to mention everyone that helped, either directly or indirectly, during the writing of this thesis. The people that made this possible are not only the ones that were around during these four and a half years of work, but the ones that made it possible for me to be here in the first place. Although I will not be able to list everyone that has made an impact over the years, there are some notable mentions.

I want to thank my supervisors, Fredrik Åström, Olof Hallonsten and Pelle Snickers for their support, without which this thesis would not be possible. You helped me shape my understanding of the fields my thesis explored, and I would always leave our discussions and meetings with enthusiasm and drive to continue forward. Thank you for the trust and encouragement to follow my line of thinking. It has been a great learning experience, but what I will remember the most, and personally hold in much higher regard, will be the trustful, friendly, and supportive environment you created. I also want to thank Katharina Cramer. Our discussions of Big Science were instrumental in figuring out my place in this interesting and broad field. Your comments on my thesis were key to helping me get my perspective across to the reader.

The external readers of my articles and kappa gave me valuable comments and advice at different points during the thesis gave me valuable advice and comments to. This includes the comments received during the Annual Swedish PhD Workshop in Information Studies with comments from Jutta Haider, Björn Hammarfelt and Gustaf Nelhans. I also want to thank the external readers for my midway seminar and final seminar, Jens Peter Andersen, and Alfredo Yegros.

The environment at the department has been extremely welcoming and nurturing. Coming from a different discipline, I had no sense of what to expect from the Department of Arts and Cultural Sciences. What I did find was open-minded and supportive colleagues that were always eager to learn from different perspectives, including mine. I want to thank everyone at the Division of ALM and Digital Cultures: Cecilia Andersson, Mikael Askander, Camila Borges Freitas, Shirley Chan, Jessica Enevold, Lisa Engström, Ola Främby, Carin Graminius, Charlotte Hagström, Charlotte Högberg, Mathias Johansson, Ann-Sofie Klareld, Sara Kärrholm, Karin Lundin, Björn Magnusson Staaf, Lisa Olsson Dahlquist, Amanda Persson, Moa Petersén, Olof Risberg, Johanna Rivano Eckerdal, Pelle Snickars, Olof Sundin, Karen Søilen, Sara Tanderup Linkis and Robert Willim. Our seminars, retreats, and social outings have made the PhD process very fulfilling. Thank you all!

I want to thank Cecilia Andersson and Olof Sundin for inviting me to collaborate with them on the interdisciplinary research project: 'Why do we choose the Internet instead of the doctor next door? The Internet as a site for medicines in grey zones.' I also want to thank the project members: Susanne Lundin, Cecilia Lenander, Rui Liu, Amelie Persson and Emma Eleonorasdotter. Being part of this project during the final stages of the PhD has been highly rewarding.

Navigating the PhD entails not only academic, but also administrative work. A lot of it. My thesis would not be possible without the support from our economist Agneta Nilsson, administrator Valeria Naters, the faculty secretary Kristina Arnrup Thorsbro, programme administrator Amanda Hindmo, director of studies Anna Burstedt, our head of department Gabriella Nilsson and all others whose work keeps the department and the university afloat.

A big thank you to all the PhD students at the department, including some visitors, and postdocs. The info-squad: Carin Graminius, Shirley Chan, and Camila Borges, for a friendly environment inside and outside of the department. Shout-out to Naoko Takayanagi, for entertaining discussions about K-pop research and for the candy from your travels; to

Raquel Garcia Plaza and David Bowling for all our friendly of just about anything. I want to thank the writing group, started by David Bowling, and now continued by Alma Aspeborg, for being the great constant in these last couple of years of the PhD, and for their support as PhD representatives.

To all the regulars at the writing group, and everyone that shared the day-to-day of PhD life. You have managed to make it better, I couldn't ask for better peers: Camila Borges Freitas, Shirley Chan, Carin Graminius, Karin Lundin, Amanda Persson, Olof Risberg, Fannie Frederikke Baden, Petra Preisler, Anni Reponen, Emma Shachat, Rikard Heberling, Magne Klasson, Anne Ladefoged, Arina Stoenescu, Molly Lindberg Alma Aspeborg, Jonas Bornsäter Lee Dallas, Simon Halberg, Josefine Löndorf Sarkez-Knudsen, Talieh Mirsalehi, David Bowling, Adam Furbring, Disa Runeby, Sarah Vorminder, Katja Heldt, Naoko Takayanagi, Phil Dodds and Brandon Farnsworth. Thank you to the PhD colleagues across the LIS departments, and other fields (some of which have since defended), with whom I have shared lectures and seminars, whose comments and support helped me keep going forward: Maximilan Weis, Nicholas Surber, Björn Ekström, Hanna Marctic, Ameera Mansour, Camilla Lindelöw, Zakayo Kjellström, Dydimus Zengenene, and the many others I have met during these years.

My friends and family have provided enormous support over the years. My friends in Sweden, we have all had great dinners, amazing hikes and travels together over the years, and I look forward to more of it. Tatiana Proisy, thank you for you enduring friendship, for believing in me and encouraging me to apply to the PhD project. To Ben Jorgensen and Sten Kjellström, for your friendship and warmth. Ludwig Schreither and Ronja Wuztler, thank you for your friendship, dinners, and game nights, which should really happen more often! Kirti John, Tamar Bzekalava, although we do not see each other often it is always a pleasure. To my oldest friends back in Mexico, Luis Gaviola, Rodrigo Herrero, Jack Levy, and Roberto Belden, thank you for almost two decades of friendship. It is always a pleasure to see you back home. To Jordy Carbonell, for bringing a part of home back to Sweden. I want to thank the Ångbäcks: Tommy, Bente, Kristian, Fredrik, and Amanda for making me feel like part of the family.

I have greatly enjoyed our time together, including dinners, parties, and travels.

I owe everything I am to my parents, Rolf Söderström, and Sofia Cozar Ayala. You have been with me in every step of the way and have supported me in more ways that I could ever say in such a short space, I am truly fortunate to be your son. Carolina Söderström, thank you for being my older sister and for the adventures you organise back home, and to Sunne, see you soon!

Finally, to Ida, I cannot thank you enough for your love and for your support during this journey, and especially these final months. We have our lives ahead of us, and I will enjoy every moment of it.

# Abstract

This thesis examines the role of beamlines with publication metadata from the European Synchrotron Radiation Facility (ESRF a large scientific facility offering diverse conditions for scientific research. As a user-oriented Big Science facility, the ESRF channels extremely brilliant light into different experimental areas called beamlines that house specific instruments that cater to the needs of diverse user communities. These beamlines are conceptualized as spaces where diverse scientific communities interact, share, and produce knowledge, drawing mainly from Derek de Solla Price's concept of instrumentalities and the works of Susan Leigh Star, Geoffrey C. Bowker, and Karen Ruhleder on infrastructure theory and boundary objects.

With the beamline as the unit of analysis, emphasising its scientific, technical, and social configurations, diverse computational techniques are applied to bibliometric metadata in four individual research articles to analyse and understand the relationship between the configuration of the beamline and their user communities. The four articles employ computational and bibliometric techniques to map the disciplinary, collaborative, geographical, and process dynamics surrounding ESRF's beamlines. The results reveal that beamlines foster collaboration and knowledge sharing among user communities. Furthermore, beamlines serve as collaborative hubs in user facilities like the ESRF, offering a space for diverse user communities to engage with in-house scientists (beamline scientists and/or technicians). While some beamlines cater to specific scientific domains, others are—or become—adaptable across disciplines.

The study emphasises that beamlines are not solely individual instruments but integral parts of a larger infrastructure for interdisciplinary research. Their technological configuration remains pivotal in shaping access and collaboration patterns amongst user communities, thus enabling specialised and interdisciplinary research. Furthermore, their evolution, including automation and remote access, have expanded the reach of beamlines.

# List of Articles

*Article I*

Söderström, K.R., Åström, F., Hallonsten, O. (2022). Generic instruments in a synchrotron radiation facility. *Quantitative Science Studies; 3* (2): 420–442. https://doi.org/10.1162/qss_a_00190

*Article II*

Söderström, K.R. (2023). The structure and dynamics of instrument collaboration networks. *Scientometrics; 128:* 3581–3600. https://doi.org/10.1007/s11192-023-04658-w

*Article III*

Söderström, K.R. (2023). Global reach, regional strength: Spatial patterns of a big science facility. *The Journal of the Association for Information Science and Technology; 74* (9): 1140-1156. https://doi.org/10.1002/asi.24811

*Article IV*

Söderström, K.R. (2023). Who did what? Creating structured data from acknowledgement texts with large language models. Submitted.

# Author's contribution to the articles

*Article I*

Kristofer Rolf Söderström: Data curation, Formal analysis, Methodology, Visualization, Writing—original draft, Writing—review and editing. Fredrik Åström: Supervision, Writing—review and editing. Olof Hallonsten: Conceptualization, Resources, Supervision, Writing—original draft.

# Abbreviations

| | |
|---|---|
| AI | Artificial Intelligence |
| API | Application programming interface |
| BM | Bending magnet |
| CERN | Conseil européen pour la Recherche nucléaire (European Organization for Nuclear Research) |
| DESY | Deutsches Elektronen-Synchrotron |
| DOI | Digital Object Identifier |
| EBS | Extremely Brilliant Source |
| ESA | European Space Agency |
| ESO | European Space Observatory |
| ESRF | European Synchrotron Radiation Facility |
| FSL | Few-Shot Learning |
| ID | Insertion Device |
| GPT | Generative pre-trained transformer |
| HHI | Herfindahl-Hirschman Index |
| ILL | Institut Laue-Langevin |
| LIS | Library and Information Studies |
| LLM | Large language model |
| LURE | Laboratoire pour l'utilisation du rayonnement électromagnétique |
| MAX | Microtron Accelerator for X-rays |
| MX | Macromolecular Crystallography |
| SESAME | Synchrotron-Light for Experimental Science and Applications in the Middle East |
| SNA | Social Network Analysis |
| SOLEIL | Source optimisée de lumière d'énergie intermédiaire du LURE |
| STS | Science and Technology Studies |
| WoS | Web of Science |

# 1. Introduction

The European Synchrotron Radiation Facility (ESRF) is a large scientific user facility that offers a diverse range of beamlines that house scientific instruments that enable a wide range of research by its users. On the one hand, the ESRF resembles Big Science like the particle physics centre CERN (European Organization for Nuclear Research), or the European Southern Observatory (ESO). These are home to major technical and scientific feats, often with long historical trajectories dating back to the European post-war period (Galison & Hevly, 1992; Weinberg, 1961, 1968). On the other hand, ESRF shares key characteristics with user facilities, a specific kind of publicly funded Big Science. They cater to thousands of scientists annually creating a constant flow of external users visiting the facility on a short-term basis (G. E. Brown Jr. et al., 2006). Access for this scientific user community is normally granted on the basis of the submission of research proposals that undergo a peer-review process, and experimental time is granted free of charge under the condition that the results are published in scientific journals (Hallonsten, 2009, 2016a). This and other conditions open the possibility for this thesis to analyse the role of beamlines at facilities like the ESRF, based on journal publication metadata and from a quantitative perspective.

The ESRF is a synchrotron radiation facility that opened to users in 1994 and completed its latest major technical upgrade in 2020, the Extremely Brilliant Source (EBS). The ESRF uses a collection of accelerators—including a linear accelerator, a booster ring, and a storage ring—to produce high-intensity X-rays for experimental research. The place where the user community works and interacts with the in-house scientists (internal beamline scientists and/or technicians) is the beamline, the experimental end-station where energy from the storage ring is directed to and transformed to serve a wide range of experimental setups.

There is a large ecosystem of these user-oriented facilities across Europe and the world, ranging from photon and neutron sources and lasers, to observatories, radio telescope arrays and biobanks. Their key purpose is to provide their user community with experimental time and specialised scientific instruments (Hallonsten, 2016a). They have a complex relationship with their user community, including an intrinsic interest in remaining attractive in terms of their capability to offer scientific and technical resources (D'Ippolito & Rüling, 2020; Hallonsten & Heinze, 2012). Furthermore, the scientific output that the users produce at these facilities in the form of journal publications often constitutes a critical indicator by which the performance of the facility is judged and assessed by policy makers and funding agencies (G. E. Brown Jr. et al., 2006; Heidler & Hallonsten, 2015).

Previous historical, political, and sociological research on user facilities highlights how these were created, are operated, and how they function within society, along with how their organisation and mission is structured by economic and political conditions, as well as societal expectations (Cramer et al., 2020, p. 202; Hallonsten, 2009, 2016a). These studies often adopt a qualitative perspective that focuses on the founding histories, organizational change, or emerging national and transnational policies. Scholarly research from the sociology of science and Science and Technology Studies (STS) also delves into the micro-level nuances of laboratory settings, aiming to understand the conditions fostering research and knowledge creation (Aarden, 2023; D'Ippolito & Rüling, 2019; Latour & Woolgar, 1986; Zemplén, 2019).

Big Science research has identified the increasingly user-oriented nature of some facilities, and their growing interconnectedness with the wider scientific, social, political, and economic system. This had led to conceptualisations such as "new", "transformed" Big Science, or research infrastructures, serving an ever larger, but also decentralized user community from varying disciplines and scientific fields (Cramer et al., 2020; Crease & Westfall, 2016; Hallonsten, 2009, 2016a; Rekers & Sandell, 2016; Westfall, 2003; Westfall & Hoddeson, 1996). Despite the important role these facilities play in the scientific user community, science, and science policymaking (Cramer et al., 2020), studies that explore impact, collaboration, and the geography of such facilities with publication metadata (Börner et al., 2021; Hallonsten, 2013, 2014; Heidler & Hallonsten, 2015; Silva et al., 2019), do not take into account the heterogenic and dynamic characteristics of user-oriented Big Science. Furthermore, a lack of a general understanding of how user communities interact with their facility (D'Ippolito & Rüling, 2019; Lauto & Valentin, 2013) complicates the analysis and assessment on these traditional questions of impact, the analysis of scientific outputs, and productivity.

Accordingly, the thesis is framed to focus on the space where this dynamic and heterogeneous characteristic of user-oriented Big Science can be explored. This space is the beamline, where the scientific user communities—mostly external to the facility make use of the unique conditions created by the facility, but refined in the beamline via scientific instrumentation, to perform their unique experimental research. The thesis investigates the role of the beamline in shaping access to the user community, including its disciplinary and collaborative dynamics around highly specialized scientific instruments with  multi-faceted, quantitative analyses of journal publication metadata at the ESRF.

The main contribution of the thesis is methodological, suggesting a collection of computational tools that arm the scientometrics toolbox with an approach to analyse beamlines and their user communities through a collection of diverse publication metadata relevant to the facility. An in-depth analysis of the ESRF is presented in this thesis, tracing connections between beamlines, in-house scientists, and the user community, quantifying patterns of scientific processes, collaboration, and knowledge sharing. It explores to what extent publication metadata can provide information about the facility through the exploration of user communities, in-house scientists, and beamlines within the ESRF.

Furthermore, the shift encapsulated by user-oriented Big Science towards providing scientific services and instrumentation to an external user community required a framework that places beamlines and their users in focus, as it is the place where they do their scientific work. Thus, the secondary contribution is theoretical, which draws mainly from Price's concept of instrumentalities (1984), boundary objects (Leigh Star, 2010; Star

& Griesemer, 1989), and infrastructure (Bowker & Star, 2000; Star & Bowker, 2010). The beamline is conceptualised as a space that shapes the access of user communities at the boundaries of different disciplinary and social dimensions through their technological configurations. The collection of which results in a complex beamline infrastructure that encompasses the facility.

The individual research articles in the thesis explore the beamlines through different methodological and theoretical lenses. Statistical, network, and geospatial analyses, in addition to natural language processing, are employed to delve into aspects of scientific collaboration, knowledge sharing, disciplinary diversity, and processes inherent to the ESRF. The relationship between those configurations and the makeup of their user communities is explored in the thesis with the conceptualisation of the beamline and with additional reports and documents from the facility. Different disciplinary, social, and geographical communities, sometimes collaborating formally, and sometimes informally, are found within these spaces, which depend to some extent on the way the beamline instruments are technologically configured. The implications of analysing such complex research infrastructures and the role of metadata in finding these dynamics are also discussed in the thesis.

## Aim, objectives, and research questions

The central premise of this thesis is that beamlines serve as the social and technological space where the in-house scientists and the user communities interact and create new knowledge and practices. Beamlines in the facility foster diverse collaboration patterns and promote interdisciplinary and inter-institutional research efforts. Thus, the two central research questions for this thesis are:

> How do beamlines within synchrotron facilities serve as unique collaborative spaces for their user communities and in-house scientists?

> How do beamline configurations and characteristics impact interdisciplinary collaboration, knowledge production, and the broader scientific ecosystem?

The aim of the thesis is to understand and characterise the unique nature of beamlines within synchrotron facilities and their role in shaping scientific collaboration and knowledge production between in-house scientists and user communities. This is accomplished by quantitatively analysing the available metadata from scientific publications in four research articles, which have the following aims:

1. To show the range and diversity of disciplinary interactions at the facility, and to demonstrate it is insightful to analyse the ESRF at the beamline level (Article I).
2. To understand how and why technological configurations affects collaboration networks in the ESRF (Article II).
3. To map the geographical distribution of the ESRF to understand its regional and global reach (Article III).
4. To discover the informal aspects of the research at the ESRF (Article IV).

Overall, the goal is to show that beamlines serve as unique collaborative spaces, and to understand how their technological configurations attract diverse user communities, influencing collaboration patterns, knowledge production, and technological development. This main goal is broken down into the two main contributions of the thesis, related specifically to the methodological and theoretical contributions.

Objective one of the thesis is to develop and validate computational techniques to extract, analyse, and map diverse metadata in the form of collaboration networks, disciplinary diversity, spatial patterns, and informal processes related to specific beamlines. The goal is to gain an understanding of how much information about the facility can be represented in publication metadata. Objective two of the thesis is to conceptualise the beamline as the main unit of analysis and show how it shapes and facilitates—i.e., configures—access for the user community. The goal is to apply this conceptualisation as the framework for studying the ESRF.

Objective one primarily concerns the methods and approaches from the four attached articles, including the use of computational methods and metadata. Although individually assessed during the peer review processes, the thesis examines how these articles when combined can enrich the understanding of user facilities. Objective two is advanced by developing the concept of the beamline in the Theory section as a space, where the collection of beamlines constitutes the beamline infrastructure. It provides the basis to re-evaluate the findings of the original research articles with this lens. Together with objective one, they provide the main points of the Discussion section.

## Sections of the thesis

The *Introduction* continues with *Placing the thesis: Big Science and information*, which aims to explore and position the thesis within Library and Information Studies and the subfields of scientometrics and bibliometrics. It provides context for the approach in the thesis and aims to clarify some of the challenges presented during the research project. The final subsection *The facility and its beamlines*, provides an overview of the ESRF. It highlights how the facility emerged, and how it and the beamlines have evolved over time in response to user demands and other external challenges. It aims to provide historical context, as well as a basic technical and operational understanding of the facility and its beamlines. The main sections of the thesis continue thereafter.

The *Literature Review* explores how the lens of Big Science, a concept that holds different meanings and connotations across fields, has been used to explore large-scale facilities for scientific research (Weinberg, 1961). It includes historically significant facilities and projects like the Manhattan Project, the Oak Ridge Laboratory, the Human Genome Project (Capshew & Rader, 1992; Galison, 1997; Galison & Hevly, 1992; Hiltzik, 2015; Hoddeson et al., 2011; Weinberg, 1961, 1968), and user-oriented facilities like the ESRF that serve large and decentralised scientific external user communities, leading to new conceptualisations such as "transformed", "new", or as "research infrastructures" (Cramer et al., 2020; Crease & Westfall, 2016; Hallonsten, 2016a, 2020a; Rekers & Sandell, 2016; Westfall, 2003). The key characteristic is that in the ESRF and other user facilities, most of the science is performed by an external scientific user community. As such, analysing and/or evaluating only the internal efforts would give only an incomplete picture of the facility in quantitative analysis of user-oriented Big Science facilities within the context of scholarly communication. (Börner et al., 2021; Florio et al., 2016; Hallonsten, 2013, 2014; Heidler & Hallonsten, 2015; Silva et al., 2019).

The *Theory* section draws on key concepts to conceptually develop beamlines as spaces with unique configurations that shape access, collaboration, and knowledge production. It guides and encapsulates the findings of the original four articles to explore dynamic relationships between beamlines and their user communities. The framing incorporates several concepts, which include instrumentalities (Joerges & Shinn, 2001; Price, 1984), boundary objects, and infrastructure (Bowker & Star, 2000; Star & Ruhleder, 1996). This conceptualises the beamline as a space with different technological characteristics that directly influence the facility, where the interaction of in-house scientists and external users influences the future of the beamline. As such, beamlines function as an endpoint for the facility to the outside scientific community, and the space where work and interactions happen. Beamlines serve as a meeting point that shapes a user community, and sometimes-different user communities, who interact with each other and with the in-house scientists, producing not only new knowledge but also new practices, serving as a microcosm for state-of-the-art research, where technology, ideas, and people interact daily.

The *Method* section summarises the data collection process and the diverse tools and techniques created and used across the research articles to analyse the versatility, collaboration networks, spatial patterns, and text data from the beamline perspective. The ESRF offers a wealth of metadata from more than 30,000 publications since 1994 via the ESRF-ILL Joint Library, pivotal for data collection. To enhance accuracy, Digital Object Identifiers (DOIs) from the library were matched with the Web of Science (WoS) database. The method incorporates aspects of data science (Blei & Smyth, 2017; Desai et al., 2022; Donoho, 2017), scientometrics, and bibliometrics to extract and analyse different sources of metadata, and to structure information from text sources. The method shows a progression from traditional statistical modelling to state-of-the-art AI-driven techniques that explores the dynamic relationship between beamline spaces and their user communities through diverse publication metadata , structured and unstructured.

The *Summary of articles* highlights the key aspects and findings of the four research articles, each exploring a different dimension of beamlines and their user communities from different metadata sources. The findings range from understanding the diversity of instruments from their disciplinary profiles (Article I) to how formal collaborations are shaped around the beamlines with a discussion of how technology plays a role in this shaping (Article II). They also uncover the geography of the user communities across Europe and the rest of the world to shed light on the reach of the facility (Article III), and some of the informal processes surrounding scientific output from acknowledgement texts. This includes mentions of technology, collaboration with other in-house scientists, and other actors (Article IV). The results show evidence for the beamline as a space for collaborative research and highlights the role they play for the user community and the facility. Furthermore, the results show a dynamic and evolving interaction between user communities and in-house scientists, sometimes physically, but sometimes with automated systems and remotely. Overall, the technological configuration of the beamline and the makeup of their user communities are shown to be interlinked.

The *Discussion* explores the role of the beamline, and how configurations of the scientific instruments of the beamline shape access and collaboration. Furthermore, the discussion covers how analysing documents and texts as informational representations of collaborative processes adds a layer that creates a more robust understanding of the beamline, the ESRF, and similar user-oriented Big Science projects. It highlights how people, technologies, and processes are intertwined in the beamline infrastructure, and how this perspective, combined with the methods, leads to more visibility of such infrastructures.

*Concluding Remarks* wraps up the thesis, exploring potential avenues for future research.

## Placing the thesis: Information and Big Science

The field of Library and Information Studies (LIS)—referred to as 'Information Studies' within this thesis, emerged from efforts to develop librarianship, bibliography, and documentation into a single, organised field (Buckland, 2012). It is concerned, amongst other things, with the information recorded in documents, with meaning and knowledge, providing unique ways of thinking about the relationship between documents, text, and information (Bawden & Robinson, 2012b). Several of its subfields provide a methodology—or a set of methodologies—to solve issues related to the type of information that is shared amongst scientists and impacts the overall stock and structure of scientific knowledge: scientific publications. On the one hand, the thesis is placed within these subfields that explore scholarly communication, which include bibliometrics, scientometrics, and/or informetrics (Bawden & Robinson, 2012a). On the other hand, the topic of Big Science is also the domain of other fields like history, sociology, and economics with their own sets of methodology, theory, definitions, and assumptions (Cramer et al., 2020; Galison & Hevly, 1992; Hallonsten, 2016a; Joerges & Shinn, 2001; Rosenberg, 1992; Westfall & Hoddeson, 1996).

The position of the thesis in this interdisciplinary space was one of the key challenges when drafting the conceptual framework and methodological approach. This is not uncommon within the field. Information studies has a history of cooperation with other disciplines, borrowing concepts and tools for areas like information retrieval from computer science and documentation techniques from the field of bibliography (Åström, 2006). This opens an opportunity to explore Big Science projects that so far have been mostly studied by historians, sociologists, or STS scholars with the potential to enrich the analysis from the perspective of information. This resulted in the need to reconcile the several perspectives of Big Science and LIS, mirroring the struggle to balance competition and cooperation with other fields, with shifting boundaries and identity with LIS (Åström, 2006). However, the central focus of studying how information is manifested remains, adapting to the challenges (Petras, 2023) surrounding the study of Big Science through scientific outputs.

The exploration of methods and metadata for the analysis of user-oriented Big Science required an inquiry into the different methods and perspectives surrounding scientific publications and documents by the similar—but still distinct—approaches to the analysis of scholarly communication, central to this thesis. One of the first challenges was unpacking the difference between these often-interchangeable terms of 'bibliometrics,' 'scientometrics,' and 'informetrics':

Bibliometrics has been defined as the quantitative study of literatures, as reflected in bibliographies, which provides evolutionary models of science, technology, and scholarship (White & McCain, 1989); or the quantitative study of physically published units, bibliographic units, or the surrogates of either (Broadus, 1987). Scientometrics was originally defined as the study of the quantitative aspects of the process of science as a communication system (Nalimov & Mulchenko, 1971), rising to prominence as the name of the journal founded by Tibor Braun in 1977 (Tague-Sutcliffe, 1992, p. 1), with some recent approaches generally more akin to the quantitative study of science and technology (Mingers & Leydesdorff, 2015; Waltman et al., 2020). Tague-Sucliffe (1992, p. 1) emphasises the role of scientometrics as the quantitative aspects of science as a discipline or economic activity, a component of the sociology of science with applications to science policy-making, where the analysis of the publication overlaps with bibliometrics. Finally, informetrics, which is the newest term of the three, was first proposed by Otto Nacke in 1979 as the measurement—via the application of mathematical methods—of wider information phenomena (Bawden & Robinson, 2012a).

Overall, bibliometrics, scientometrics and informetrics are fields associated with the study of the dynamics of disciplines, as is reflected in the production of their literature, with similar and overlapping methodologies. As such, there has been extensive discussion over the terminology of these three closely related metric terms (Brookes, 1990). Hood and Wilson (2001) claimed that this was a result of a failure to recognise that there is more to science than its output of literature.

Since its development as a field, scientometrics has become a major player in the social and political processes of the academic community due to the drive of governments and other bodies to monitor, record, and evaluate the performance of research and researchers (Mingers & Leydesdorff, 2015). It has also received wide attention in recent decades from other fields for its applicability in the analysis of scholarly communication across diverse fields and for its potential to successfully model the dynamic and multifaceted science system with computation-intensive models in a multidisciplinary approach (Lucio-Arias & Scharnhorst, 2012).

A second challenge was to unpack the different perspectives to the approach, methodological and theorical, in the subfields. The relational and citation paradigm in bibliometrics has remained dominant, as the act of relating to other research provides a link between people, ideas, journals, and institutions, constituting empirical fields or networks that can be analysed quantitatively and traced over time (Mingers & Leydesdorff, 2015). This perspective has evolved beyond the citation as the sole focus, with refinements exploring methods and approaches like co-citation, co-word analysis, and analysis of collaboration through co-authorship (Callon et al., 1991; Small, 1973).

Furthermore, others have explored new avenues for research, enabled by the web and by computation (Cronin, 2014). Areas like science mapping (for a review of the literature and tools see Chen, 2017; Moral-Munoz et al., 2019); spatial scientometrics (Frenken et al., 2009; Frenken & Hoekman, 2014; Gao, 2020), and the analysis of acknowledgements text (A. Díaz-Faes & Bordons, 2017; Álvarez-Bornstein & Montesi, 2020; Paul-Hus & Desrochers, 2019), are some examples of avenues more analogous to an exploration of the representation of processes of empirical phenomena in documents and their metadata.

Bibliometrics and scientometrics are well acquainted with theoretical grounding. The fields have historically used mathematical models to achieve a systematic understanding of the structure and evolution of scientific knowledge, and in fact, they have influenced a considerable number of scientific disciplines. Three main models within information studies remain influential: Lotka's (1926) frequency of publications law, Goffman and Newill's (1964) epidemic model, and the network model introduced by Price (1965). Some viewed the role of theory as crucial for scientometrics to remain a scientific discipline, rather than a technological community of researchers applying standardised methods (Gläser & Laudel, 2001). Furthermore, the influence of network science, statistical physics, and complexity research to scientometrics, show the adaptation of scientometrics to other perspectives, creating and combining concepts and ideas within and between fields (Lucio-Arias & Scharnhorst, 2012).

Generally, more and richer data are becoming available, some fading into the background as nearly all electronic devices use it or generate it (Pomerantz, 2015a). Bibliometric metadata are no exception to this trend, as more granular publication metadata (Alemu & Stevens, 2015b) are being made available over time. Furthermore, as metadata, computing power, and algorithms become widespread and adopted by researchers, the question as to whether qualitative and quantitative approaches to science studies should be integrated deserves another look. Indeed, there is interest in filling in the gap between fields to develop new tools that make use of the richness of the recent data explosion, which allows efforts to shift from simulating to mapping, and from simple explanations to complex observations (Venturini et al., 2015).

There is a potential for integrating theories and insights from qualitative meta-sciences that could help craft a better understanding in the quantitative fields of scientific studies (Heinze & Jappe, 2020; Leydesdorff et al., 2020; Waltman et al., 2020). Documents—and in this case scientific publications—provide rich—structured and unstructured—metadata sources, with different levels of readiness to explore new perspectives (Pomerantz, 2015b), with the potential to analyse their statistical properties to understand of the social processes around information (Bates, 2005). Furthermore, calls to promote fair and transparent use of publication metrics for research and evaluation (Hicks et al., 2015), and the view of scientometrics as an ecology where different types of assessment procedures, funding arrangements, and allocation systems interact (Hammarfelt & Åström, 2020), have been influential in the empirical approach adopted in this thesis where methods are infused with a theoretical perspective to enrich analyses. The study of infrastructures (Bowker, 1994; Bowker & Star, 2000; Star & Bowker, 2010; Star & Ruhleder, 1996), which have been influential in LIS (e.g., Andersson & Sundin, 2023; Sundin et al., 2023; Sundin & Haider, 2022), and STS (e.g., Sawyer et al., 2019; Vertesi, 2019), to name few, offer a fruitful perspective to study how technology influences the social organization of science (Fecher et al., 2021), central for this thesis.

## The facility and its beamlines

The ESRF originated in the late 1970s in France, and political support for the facility matured in the early 1980s with additional backing from Germany. While some tensions arose during the initial tripartite negotiations with the United Kingdom, , the facility was eventually established in 1988 as a limited liability company under French domestic law by the agreement of 11 European countries. In 1994, construction was ready, and the facility was opened to the user community for experimental research (Cramer, 2017). Over the next few decades, the ESRF would cement itself as a pioneer in synchrotron radiation research, attracting a bigger and more diverse user community over time. Currently, the ESRF is comprised of 21 partner countries. It currently hosts around 44 beamlines that simultaneously serve different experimental opportunities to a multidisciplinary community (ESRF Highlights 2022, 2023). In 2020, the ESRF launched the Extremely Brilliant Source (EBS), a new high-energy light source that increased the capabilities of the facility and its beamlines, meaning higher energy levels and ranges that can accommodate more kinds of experimentation (ESRF Highlights 2022, 2023; Raimondi, 2016).

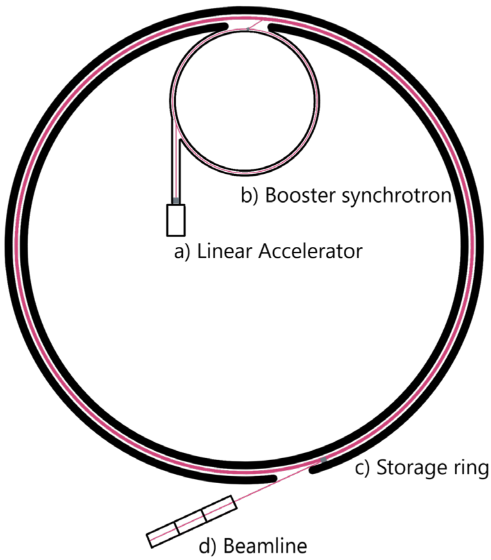


**Figure 1.** Simplified schematic of the main components of ESRF, including the a) linear accelerator, the b) booster synchrotron, and the c) storage ring, which has a circumference of around 800 metres. Surrounding the storage ring is a collection of over 40 d) beamlines. The coloured line illustrates the trajectory of the beam, which originates from the a) linear accelerator. Within different sections of the storage ring, magnets are used to direct and focus the energy towards the beamline. Typically, a beamline comprises three distinct areas: an optics cabin, an experimental cabin, and a control cabin. The schematic was created by the author.

Figure 1 shows a simplified schematic of the ESRF. Electrons are launched by an electron gun at the start of the linear accelerator (a), accelerating them close to speed of light. They then enter the booster synchrotron (b), travelling around it several thousand times, getting more energy with each lap. The electrons are then deposited in the storage ring (d) where they can travel for hours at the near the speed of light under extreme vacuum conditions. Electrons pass by several types of magnets, losing energy in the form of electromagnetic radiation or "synchrotron radiation." This radiation enters the beamline (d), where the user experiments take place. The experimental work by the user community is carried out in these collection of relatively small spaces around the storage ring, equipped with state-of-the-art-instrumentation (European Synchrotron Radiation Facility, n.d.-b, n.d.-d).

Very simply, the beamline is the device that accomplishes the transport of a charged particle from the accelerator to an end station (Dehnel, 2011). However, the definition of the beamline by the ESRF has shown some variety, which provides insight into the importance in understanding the role of the beamline in the facility, beyond that of a simple delivery device. At the early stages of the thesis, three definitions by the ESRF were found: first, a cylindrical pipe that directs the beam generated in the storage ring to the experimental end-station; second, the scientific instruments in the experimental end-station; and third, the dynamic meeting place where teams do research (Article 1) .

Since then, it has also been defined by the facility as highly specialised laboratories with state-of-the-art instruments with dedicated support teams (European Synchrotron Radiation Facility, n.d.-a, p. 5). Beamlines house these scientific instruments in different areas, or cabins: the optics cabin, the experimental cabin, and the control cabin. The optics cabin is the first in the path from the storage ring and gives the beam the necessary characteristics for the type of experiment being conducted. The experimental cabin contains the positioning device that handles the sample to be studied, with detectors that record data generated by the beam as it reaches the sample. Finally, the control cabin is where the scientists control their experiments and collect the data, with remote capabilities in some beamlines (European Synchrotron Radiation Facility, n.d.-h).

The definition of the beamline will be revisited in the Theory section. For now, what remains relevant is that the beamline is where experiments take place—equipped with state-of-the-art scientific instrumentation and a dedicated team of in-house scientists that not only build and upgrade the beamline, but also support its users whenever needed.

The ESRF has seen constant and periodical approval and development of upgrades to the storage rings and beamlines throughout the past few decades. This focus on change and adaptation is crucial to secure operation and funding for such facilities and projects over the long run and has been explored in previous research, where technical, political, and social aspects have been brought into focus (D'Ippolito & Rüling, 2019, 2020; Heinze & Hallonsten, 2017; Westfall, 2008). The evolution of ESRF's technological capacities and capabilities not only encapsulates automation in sample screening and data collection but also the introduction of digital tools such as beamline databases for efficient sample tracking, experiment reporting, and real-time remote monitoring (Malbet-Monaco et al., 2013)

Support for ESRF's approximately 5,000 external users per year—which has grown closer to 9,000 in the latest years (ESRF Highlights 2022, 2023)—can come in many forms, including face-to-face communication, and learning whenever a user visits the experimental station to perform an analysis on a sample. However, the support can also take the form of enabling automated routines and remote access with physical instruments and/or developing software for data analysis. The use and support are free of charge for the scientific user community, which go through a peer-review process to use the beamline in the facility (Hallonsten, 2016a; Hallonsten & Heinze, 2015).

One important consequence of this close connection between the in-house scientists and the external users is that the ESRF and similar facilities depend on incorporating the innovative demands of their users to remain attractive. Hallonsten (2009) points out that experiments require, but also establish close engagements and ties between the external users and in-house scientists at the facility. This translates into the development of instruments and experimental settings and shapes the generation and diffusion of innovative knowledge and the formation of collaborative ties among researchers.

Automatization processes have been a major technological concern for the ESRF in the most recent decades. For example, the Medium and Long-Term Scientific Programmes (Bouvet et al., 2007; European Synchrotron Radiation Facility, n.d.-g, n.d.-f), which provides an overview of the financial and technological planning for the 2000s. They deemed automation a necessary development for the facility, seeing it as a principal means of keeping the facility in operation and providing enhanced support to users (European Synchrotron Radiation Facility, n.d.-g). Although it would require considerable financial and human resources, it would pay off in the long run, leading to new and exciting challenges in synchrotron radiation research, enabling in-house scientists to remain competent in their research programmes, and enabling emerging scientific applications of synchrotron radiation (Bouvet et al., 2007; European Synchrotron Radiation Facility, n.d.-g).

Automation in beamlines within fields of Structural Biology were already in development at that time, showing the focus on automating beamlines both in terms of the physical manipulation of samples, as well as the collection and analysis of data (ESRF Highlights 2002, 2003; ESRF Highlights 2004, 2005). By 2006, some beamlines had already incorporated remote access capabilities into their automatic data collection systems, allowing users to remotely interact with beamline control systems, facilitating automatic sampling, data collection, and processing through intricate modules and web services (Beteva et al., 2006), to facilitate time-intensive routine tasks.

Macromolecular crystallography beamlines, often used in Structural Biology and part of the broader set of beamlines in the ESRF, have consistently trended towards more advanced automation and digitalisation since the early 2000s, including increased data collection speeds and higher sample throughput rates, aiming to enhance user experience (Gabadinho et al., 2008).

Furthermore, a substantial increase in remote access was witnessed during the global health crisis of COVID-19 (ESRF Highlights 2022, 2023). The measures adopted due to COVID-19, encompassing social distancing and travel restrictions, necessitated innovative solutions to maintain facility operations and research continuity. To adapt, ESRF instituted multiple measures, including remote access experiments and refined sample mail-in systems, which succeeded in attracting remote users worldwide through their remote interaction tools, and improved the user-friendliness of their environment (Chenevier et al., 2021).

# 2. Literature review

Each research article in the thesis contains the relevant literature review for the specific metadata dimension analysed for the ESRF, and in some cases, overlap will be found with this section. However, the main aim of this literature review is to provide an overview of the diverse ways the ESRF and similar facilities have been analysed over time, with conceptual frames like Big Science, new or transformed Big Science and as research infrastructures. It also explores how quantitative approaches have explored these facilities with the use of publication metadata.

## User-oriented Big Science

The term 'Big Science' has a dual and interesting history and roots the thesis to two diverging paths in research. The two understandings of Big Science were originally brought to the public light independently by two influential scientists in the 1960s, although with slightly different perspectives in mind and reaching different audiences. On one hand, it refers to large-scale facilities and projects requiring big funding, organisation, and machines (Cramer et al., 2020; Hallonsten, 2016a). This is first attributed to Alvin Weinberg (1961, 1968), then the administrator of the Oak Ridge National Laboratory in Tennessee, US. On the other hand, it is related to science and technology in general (Cramer et al., 2020), and specifically to the exponential growth of scientific work and publications first identified by Derek J. de Solla Price (1963), largely regarded as the driver behind the creation of scientometrics as a field of study (For a deeper analysis see Capshew & Rader, 1992).

Price (1963) saw Big Science as an unsustainable interlude of how science is done, from a solitary "little science" to a largely team-driven enterprise.

Weinberg (1961) advised against the over-reliance of large-scale government-funded facilities that focused solely on high energy particle physics, as its unprecedented growth would lead to bureaucracy, and the loss of academic freedom and serendipity in science. This critique was also related to how other domains in science could benefit from the resources that were taken up by his own field, such as medicine or environmental sciences, saying:

> "I personally would much rather choose scientific issues which have more bearing on the world that is part of man's everyday environment, and more bearing on man's welfare, than either high-energy physics or manned space travel." (Weinberg, 1961, p. 164).

The study of Big Science facilities and projects matured as a field of study in the 1990s with seminal works by Capshew & Rader (1992) and Galison and Hevly (1992), during which Big Science or 'big machines', 'big organisations', and 'big politics', stem from the Cold War era of fundamental physics research driven by symbolic and cultural capital (Hallonsten, 2016a). These are facilities, or projects with a single purpose, with mostly in-house scientists. In fact, this is what most people—both the wider academic and public spheres—conjure up in their minds when they hear Big Science, ushered by the Manhattan Project, and cemented with facilities like the Large Hadron Collider at CERN and projects like the Human Genome Project (Capshew & Rader, 1992; Galison & Hevly, 1992; Weinberg, 1968). However, questions surrounding the emphasis on the 'bigness' of Big Science would soon emerge.

Westfall and Hoddeson (1996) explored the development of the Fermilab particle accelerator, which faced budget constraints that demanded flexibility and adaptability from their scientists due to limited funding. Galison (1997) introduced the concept of 'trading zones' to explore the need for teamwork in operating multimillion-dollar machines, where instrument makers, theorists, and experimentalists meet, share knowledge, and coordinate the diverse fragments of modern microphysics.

Westfall (2003) continued attempts to rethink the concept of Big Science, urging academics to look beyond high-energy particle physics laboratories to better understand the dynamics of large-scale research. To this end, Westfall analysed the development of the Bevalac, a modestly sized science facility that was built from two existing machines: the Super-HILAC and the Bevatron. The new machine received a lukewarm response as it did not fit the needs of the incumbent high-energy physics communities. However, it was particularly useful for nuclear physicists, opening new possibilities of research and experimentation for that user community (Westfall, 2003).

Crease and Westfall identified the changes from “old” to “new” Big Science in the United States, where fields like Materials Science, which look at much larger scales than the traditional subatomic scales within physics were emerging in facilities both smaller and more complex. This “new” Big Science was more accountable to funders, developed more practical applications, had more industrial involvement, and had a more diverse user community (Crease & Westfall, 2016).

In fact, the X-ray energy created at synchrotron radiation facilities like the ESRF, once deemed ‘parasitic’ by its creators due to the energy ‘leakage’ during the particle acceleration process, showed potential for other science and research avenues. Some scientists saw the potential and started new types of experimentation (Hallonsten, 2015). This transformation eventually led to the creation of new facilities, purposely built to produce this type of energy: radiation from accelerating and bending electrons close to the speed of light. A new type of organisation had emerged, and Weinberg’s wish for a multidisciplinary Big Science was granted. This contemporary form of Big Science— ‘new’ or ‘transformed’—is multidisciplinary, mostly performed by research groups from other universities, institutes, and industry, undertaking daily experimental work on specialized scientific instruments (Cramer et al., 2020; Crease & Westfall, 2016; Hallonsten, 2016a).

In contrast to the “old” Big Science facilities, these facilities host not only their own research, experimentation, and scientists but are also host to a larger external multidisciplinary user community which visits the facility to perform experiments as part of their ordinary scientific work (Hallonsten, 2016b). The user community performs their own research with the help of different instruments available for experimentation within the facility, without a centrally planned objective or collective goal by facility managers. The facility, in turn, aims to help solve current challenges developing around fields like materials and health sciences with increased accountability from policy-makers and the public (Cramer, 2017; Hallonsten, 2014, 2016a, 2016b; Heidler & Hallonsten, 2015).

The differentiations between the in-house scientists and the external user community, the presence of 'normal' or 'small' science, the highly multidisciplinary work done at the facility, and the role within the larger scientific, social, and economic systems have also led to the conceptualisations of user facilities like the ESRF as a 'research infrastructure'(Cramer et al., 2020; Hallonsten et al., 2020). The term research infrastructure is a relatively new conceptual framework, and like Big Science, it appears to have dual interpretations depending on context. One within organisational studies and the other within a European policy context. The former, non-capitalised research infrastructures are defined as comparably large and centralised physical- and technically advanced resources used for experimental research in the natural sciences. The latter, 'Research Infrastructures' capitalised, have emerged as a policy-oriented concept within European policy frameworks. The former definition includes a wide collection of entities, including large scale science facilities, but also physically distributed or decentralised resources for research, such as computing networks, big data, physical objects, and user communities (Cramer et al., 2020).

Framing user facilities as research infrastructures highlights their functionally differentiated role as stable and durable resources and enablers of scientific research. They are a physical and resilient space for breakthrough research, providing services of high quality to scientific communities, growing into a fundamental part of the innovation system, where reliability and quality are bases for collaborative activities within the system. As different components of the system contribute to this goal, it is necessary to identify and separate their functions (Hallonsten et al., 2020).

The overlap between the three conceptualisations of user facilities—"new," "transformed," and as a "research infrastructure"—reflects both the ongoing internal and external changes to the facility, as well as attempts to contextualise and understand them. The European Strategy

Forum on Research Infrastructures (2018) states that research infrastructures have the potential to offer an environment that generates a high flux of peer-reviewed proposals and experiments that stimulate international collaborations, where several scientific disciplines and economic sectors cross together: physics, chemistry, biology, Earth sciences, energy, cultural heritage, food, etc. They provide instrumentation, or the ability to develop new instruments to find solutions to demands (European Strategy Forum on Research Infrastructures, 2018).They offer opportunities and services that play a key role in the performance, innovative strength, and international competitiveness of science, playing an important part in generating new knowledge and technologies. Their strategic importance and funding need result in growing demand for assessing their scientific output and impact (Fabre et al., 2021).

Regardless of conceptualization, understanding the relationship between the facilities, including its physical infrastructure, and their user communities, heterogeneous, dynamic, and extensive, is crucial. As the facility has no centrally planned objective or collective goal (Hallonsten, 2016a), but instead are:

> "created and sustained by the interaction of two primary forces: the unification provided by the central physical infrastructure (the accelerator) and its operation, and the disunification of the dynamic and varied scientific program, consisting mainly of external users who come and go" (Hallonsten, 2016a, p. 108).

The dynamics between the user community and the facility is key to understanding its role as a research infrastructure in a wider ecosystem (Hallonsten et al., 2020). Some facilities have been explored from this perspective, which highlight the advantages for exploring the user community. For instance, D'Ippolito & Rüling (2019) find four typical collaboration patterns, which reflect the perceived expertise gap between beamline scientists and users and co-development related to the instruments: full-service, complementary, instrument service, and peer collaboration, whose findings suggest that a wide range of collaboration types, are beneficial for the facility. However, a general understanding of how user communities interact with their facility remains scarce (D'Ippolito & Rüling, 2019; Lauto & Valentin, 2013), despite the important role these facilities play in science and science policymaking (Cramer et al., 2020).

## Quantitative studies of Big Science publications

Big Science, as conceptualised by Price (Capshew & Rader, 1992; Price, 1963, 1965), together with the methods of bibliographic coupling (Kessler, 1963) and citation indexing (Garfield, 1955) helped establish the scientometrics field. As explored in the Placement of the Thesis, this quantitative approach utilises the abundant publication metadata sources for large-scale analysis, keeping up with the growth of scholarly communication. The approach has been adopted to analyse and evaluate subjects, projects, universities, and Big Science facilities. Despite advancements in the field (Cronin & Sugimoto, 2014; Hicks et al., 2015), output and evaluative bibliometric measures like publication are often the focus of such analyses, and directly associated with scientific performance by lab directors, advocates, and policymakers (Abramo & D'Angelo, 2023; Wouters, 2014). However, the heterogeneity and dynamism of user-oriented Big Science exposed above, including the lack of a centralised structure to analyse and evaluate, complicate assessments on these more traditional questions of impact by user communities, or analysis of scientific outputs, and productivity.

The study of Big Science facilities using a bibliometric or scientometric approach is scarce. In initial attempts to tackle the problem quantitatively, Hallonsten (2013) introduced and explored "facilitymetrics", in which technical reliability, competition for access, and publication records are assessed as quantitative performance measures. Despite being used and propagated by facilities, they are deemed insufficient for assessing their impact and productivity. For instance, crucial factors like the quality of user support or the reliability of lab instruments—crucial for the user community—cannot be accounted for in this approach. As argued and developed further by Hallonsten (2014), simple measurements like publication counts are not appropriate for analysing the complexities surrounding these facilities, especially when comparing them to expenditure figures. Although the approach is intentionally simplistic, it highlights the challenges of using cost-per-publication as a metric for success in these facilities, suggesting the need for other methods for assessing success. Heidler and Hallonsten (2015) expanded "facilitymetrics" by adding the dimension of impact and exploring metrics like the impact factor (Garfield, 1955), immediacy index, and citation network analysis. They find that these are still not enough to capture the unique capabilities of the facilities and contributions to science; for instance, the ability to stimulate and sustain fields at the intersections of traditional scientific disciplines.

Silva et al. (2019) provided a quantitative assessment of user facilities by combining network and citation analysis. The authors build a co-authorship network of internal (in-house) and external scientists associated with the facilities, with careful consideration of the limitations of data collection mentioned earlier. Their findings indicate a similarity in research quality between internal and external users, and argue that, for simplicity, one could assess the quality of a facility based solely on its internal users. However, over 60% of the publications are identified as external to the facility, which come to the facilities to solve problems that are more multidisciplinary in nature than the ones faced by in-house scientists (Hallonsten, 2016b). Omitting external users from any study would be detrimental to explore the facility and how it relates to its multidisciplinary user communities. Furthermore, an analysis on internal users would erase the networks of user communities formed between in-house and external users, reinforcing the view of these facilities serving a narrower set of disciplines, akin to "old" Big Science facilities.

Due to the significant presence and constant influx of external users, user-oriented Big Science facilities cannot be evaluated in the same manner as typical scientific units such as research groups, institutes, departments, schools, or universities (Hallonsten, 2016a). There exists a dynamic interplay between users and facilities, where users often seek the best opportunities, sometimes on a global scale, to access the technologies essential for their work (Hallonsten, 2016b; Joerges & Shinn, 2001). Therefore, the performance of the facility is not tied necessarily to its productivity, and it is up to the user communities to deliver on basis of their skills, competences and choice of research topics and problems (Heidler & Hallonsten, 2015).

What is overlooked in the scarce quantitative literature surrounding these facilities is the exploration of the interaction between the user communities and the facility. Focusing on the beamlines within ESRF—one can instead focus on how these allow user communities and in-house scientists to interact. The following section serves as the building blocks that frame the analysis. It provides a theoretical basis to the reduction of the unit of analysis from the facility to the beamline. It aims to understand what this space is, how it is configured technologically, and how this affects the production of knowledge, practices, and social aspects.

# 3. Theory

The basis for focusing on the beamlines is evidenced by both the wide disciplinary range, as well as the attention to the external users of user-oriented Big Science, explored in the previous sections. As previously detailed, the beamline is the location where most user science occurs, and facilities extensively document their status, changes, and upgrades. They also organise their information from the beamline perspective. They record beamline information on the website, in reports, in outward communication and press conferences, and their publication library (EPN-Campus, 2014). Furthermore, the shift encapsulated by user-oriented Big Science facilities towards providing scientific services and instrumentation to an external user community requires a framework that places beamlines and their users in focus, as it is the place where they do their scientific work. Therefore, this section scaffolds the conceptual framework that assists in guiding and interpreting the results from the research articles. As a compilation thesis, some of these concepts have been developed for the individual articles, acting as building blocks for the larger framing in the thesis.

Overall, the framing locates the analysis in the space within the facility where teams of scientists external to the facility do their scientific work. The beamline, where they interact and share knowledge with each other and with other teams, as well as with in-house scientists, is facilitated by the scientific instrumentation available within it.. This framing is based on the knowledge gained from writing each article, undergoing peer review, and participating in seminars and coursework throughout the project. Thus, some changes in the framing between the first and last articles are evident and will be explored in the *General discussion*.

In the case of the ESRF, an analysis based on a top-down aggregation of the facility—in which all beamlines are treated equally and aggregated into a coherent framework—becomes too complex a task, making the role of the beamlines, and their dynamic relationships with the user communities invisible. Furthermore, beamlines have unique scientific instruments that allow them to adapt to different user communities and their needs. Thus, socio-technical and complexity (Benbya et al., 2020; Leonardi, 2012; Trist & Bamforth, 1951; Vemuri, 1978) perspectives form the foundation of the standpoint adopted by the thesis. Although these are echoed in the theories and concepts this section draws upon, placing it in these settings offers a foundation for the rest of this section, serving as a baseline for studying the facility, the interactions of the user community and the in-house scientists at the beamline level, where people, processes, and technology come into play. The aggregate beamline infrastructure is instead analysed from a bottom-up perspective. This observation provides the starting point for understanding of the relationship between scientific instruments, beamlines, facilities, and user communities at differing levels of aggregation. It also helps to consider what success in a decentralised research infrastructure looks like, and how facilities adapt to a changing environment and to changes in the needs of the user community. The next step in the scaffolding revolves around understanding what constitutes the beamline beyond a set of individual scientific instruments. Hence the role of the beamline is explored, at the intersection of processes, people, and technology.

## The beamline

The fluid definitions of the beamline by the ESRF deemed a more careful exploration for the thesis. Thus, definitions from a collection of leading synchrotron radiation facilities are collected to provide a more robust picture of the beamline, its scientific instruments, and their function in the facility:

The French national synchrotron facility SOLEIL defines the beamline as the experimental facilities of a synchrotron laboratory (SOLEIL Synchrotron, 2016). Adding:

> "They consist of one or several successive cabins where the beam is propagated from the storage ring, selected, focused, and directed toward the samples being studied. Each beamline is specialised and equipped for one or a few techniques of analysis." (SOLEIL Synchrotron, 2016).

The Synchrotron-light for Experimental Science and Applications in the Middle East (SESAME) provides its own definition:

> "…light is collected by different 'beamlines' connected to the storage ring: beamlines contain the optical elements that select and focus certain wavelengths of the synchrotron light on materials that scientists wish to study, as well as the set up for controlling the sample's environment and for data collection. Each beamline is designed to produce light with characteristics that are suited for a specific type of research. […] They are also the physical area within the experimental hall where the scientists visiting SESAME, referred to as 'users', carry out their experiments. They are the work place where the users, often with diverse cultural, political, and religious backgrounds, interact on scientific issues and through this build cross-border scientific collaboration, dialogue, and understanding." (SESAME, 2020).

The British national synchrotron facility, the Diamond Light Source, provides the following definition:

> "Synchrotron light is emitted when a beam of electrons moving close to the speed of light is bent by a powerful magnetic field. The light that is produced spans the electromagnetic spectrum from infrared, through visible and ultra-violet light to X-rays."[…] "Beamlines typically have three hutches; control cabin, experimental hutch, optics hutch." (Diamond Light Source, 2022).

Finally, the Swedish national synchrotron facility, MAX[1] IV, defines the beamline as:

1 Microtron Accelerator for X-rays

> "…providing modern X-ray tools to address scientific questions with spectroscopy, diffraction, and scattering, or imaging techniques." (Rift, 2023).

It is possible to find common elements to construct a working definition for the thesis: The beamline is an integral component of a synchrotron radiation facility that, given a set of specialised instruments, is responsible for collecting, selecting, focusing, and directing the synchrotron light emitted when a beam of electrons, moving near the speed of light, is bent by a magnetic field in the storage ring. This light spans the electromagnetic spectrum from infrared to X-rays. The scientific instruments tailor specific wavelengths of this light for varied and detailed studies. Beamlines often consist of specialised cabins or hutches, including control, experimental, and/or optics sections, which manage energy levels and ranges, and with additional scientific instrumentation for sample handling, data collection, and analysis. Each beamline is designed for research techniques and serves as the primary location where scientists—both internal and external to the facility, and often from diverse backgrounds and disciplines—conduct experiments and foster collaborative science.

## Beamline space

During the early stages of the thesis, beamlines were considered solely as scientific instruments, and as such, the concept of generic instruments (Joerges & Shinn, 2001; Rosenberg, 1992) was used to explore the role beamlines play as tools that enable scientific research across a wide variety of disciplines and social practices . Scientific instrumentation, in more general terms, technology, has been shown to be enablers of new scientific knowledge and practice, with new avenues for research that solidifies theoretical findings in science (Shrum et al., 2007). The exploration of the role that these instruments play in science was one of Derek J. de Solla Price's many influential contributions. Price argued:

> "I advocate the use of the term instrumentality to carry the general connotation of a laboratory method for doing something to nature or to the data in hand" […] "A common feature of instrumentalities is that they are

> rarely accorded full recognition at birth: almost nothing would lead one to predict that a given technique would yield decisive results. One might never expect that an improvement in spectacle lens-grinding would change astronomical cosmology." (Price, 1984, p. 13).

Price further emphasises the serendipitous nature of discoveries made possible by new instrumentalities. In a lab setting, these tools can lead to unanticipated phenomena, offering the potential for significant breakthroughs. Similarly, when introduced to the commercial market, the right strategies can allow the same tool to address an existing or even unidentified need, creating opportunities for new applications (Price, 1984). However, contrary to the instruments in beamlines, Price (1984) considered them as inward-looking, single-domain technologies, not necessarily connecting to other disciplines. However, as the results of Article II will show, the ESRF displays cross-and-multi-disciplinary interactions along the beamlines that are shaping collaboration, affecting how knowledge is shared, and informing upgrades with the potential to be useful in other disciplines. Furthermore, as work for the second article was concluding, it seemed clear that the beamline was not solely a generic instrument in isolation, as it influenced how knowledge and ideas spread between user communities (Burt, 2004; Granovetter, 1973) .

Instead, the beamline appears as a suite of scientific instruments that make up the technological characteristics that influence and make the work of these user communities possible. These are spaces, or 'trading zones' as described by Galison (1997), where teamwork is fostered; instrument makers, theorists, and experimentalists meet, share knowledge, and coordinate, making it possible for teams to share a local understanding of a complex topic via agents (Galison, 2010), or in-house scientists physically placed within the beamline space, housed in different hutches (as explored in The facility and its beamlines).

Moreover, scientific instruments neither remain static over time nor remain in isolation from the user communities they serve. Instead, positive feedback loops (Arthur, 1990) between beamlines and user communities emerge. The development of scientific instruments is a process that continually changes the beamline spaces, a combinatorial exercise, where new technology is created on top of existing technology, which over time leads to the creation of entirely new ways of interaction and knowledge (Arthur, 2009). This process of innovation is incremental, and involves problem solving, adaptation and combination of elements in novel ways. Thus, the technological configuration of these instruments results in different beamline spaces over time that can adapt to different user communities with their unique challenges.

The combinatorial and adaptive characteristics of the beamline imply that each will have different configurations of scientific instruments, which will determine the way the beamline is used, which users come to it, and its outcome or impact. Recognising how it shapes access to user communities, and how they converge, is crucial. This perspective also paves the way for a systematic understanding of the user community based on the unique beamline configurations, and their relationship. Not only do the configuration of the beamline space influences the user community, but the user community influences the beamline configurations in an incremental feedback loop (Arthur, 1990, 2009)

This conceptualization would imply that the configuration of different scientific instruments within a beamline attracts specific user communities, each with their unique disciplines and practices, but also bridge gaps between different user communities, facilitating collaboration and knowledge transfer between them. The generic (Joerges & Shinn, 2001; Rosenberg, 1992) nature of some of the scientific instruments in the beamline allows for fine-tuning beyond the original designs, based on the evolving needs of current and newer communities. The collection of beamline spaces is embedded within the synchrotron facility and is essential for its functioning as a user-oriented facility, and the collection of beamlines would entail an infrastructure that enables research beyond the simple physical aggregation of the beamlines in the facility.

## Beamline infrastructure

The concept of infrastructure (Bowker & Star, 2000; Star & Bowker, 2010) has been applied to a wide range of issues within the context of Information Studies and Science and Technology Studies. These studies address, simultaneously, the technical, social, and organisational facets of infrastructures in global and local arenas, including their development, usage, and maintenance (Fecher et al., 2021). In terms of large-scale projects, they support research practices through an array of digital services and resources. Information infrastructures specifically enable new forms of sociality shaped by Information and Communication Technologies (ICT), which change the nature of knowledge work (Bowker et al., 2010).

Similarly, the ESRF provides, through its collection of beamlines, spaces that serve multiple communities simultaneously; balance standardisation and local flexibility; and provide objects that have local utility but shared meaning across communities (Bowker & Star, 2000; Leigh Star, 2010; Star & Bowker, 2010; Star & Ruhleder, 1996). More explicitly, beamlines function to some degree as boundary objects, which are:

> "…objects which are both plastic enough to adapt to local needs and constraints of the several parties employing them, yet robust enough to maintain a common identity across sites. They are weakly structured in common use, and become strongly structured in individual-site use. They may be abstract or concrete. They have different meanings in different social worlds but their structure is common enough to more than one world to make them recognizable, a means of translation. The creation and management of boundary objects is key in developing and maintaining coherence across intersecting social worlds." (Star & Griesemer, 1989, p. 393).

Furthermore, as Bowker and Star (2000) identified, there are situations where an infrastructure serves various communities of practice, either within a single or multiple organisations:

> "… what we gain with the concept of boundary infrastructure over the more traditional unitary vision of infrastructures is the explicit recognition of the differing constitution of information objects within the diverse communities of practice that share a given infrastructure" (Bowker & Star, 2000, p. 314).

This property of the beamline allows it to cross disciplinary and practice boundaries, through abstract or concrete objects, including the digital technologies in the beamline. The beamline infrastructure is the bottom-up collection of beamline spaces, shaped by both its technical capabilities and by social interaction. It is a collection of active scientific tools shapes the scientific process (Price, 1984), which—while designed and built for some specific purpose by beamline scientists—can sometimes be applied to and used for other purposes (Joerges & Shinn, 2001; Rosenberg, 1992). It is not only the physical space that allow these user communities to interact, share, and produce science (Hallonsten, 2020a; Hallonsten et al., 2020, 2020) but also the shared languages, practices, and digital technologies tying these beamlines together, formed by the interaction of people, software, and hardware (Åström, 2016; Bowker & Star, 2000; Star & Ruhleder, 1996). In other words, the tangible machines and people are tied to intangible practices, protocols, and norms.

An interesting duality emerges when studying these beamlines as part of a larger facility from an infrastructural perspective. The wide collection of entities, including large scale science facilities, but also physically distributed or decentralised resources for research, such as computing networks, big data, physical objects, and user communities (Cramer et al., 2020), might initially seem at odds with the infrastructure of STS and LIS, which tends to be invisible, working in the background (Bowker & Star, 2000; Star & Ruhleder, 1996) until it is noticed, often the result of some type of failure, where works of infrastructural inversion (Bowker, 1994) are common.

However, if we think of the conceptual definition of infrastructure, some parallels seem evident regarding the ESRF: a large-scale, interconnected, and standardised sociotechnical system that allows for some situated flexibility and embedded, working systems that support different communities, and an interdependence between their technical networks, i.e., the collection of interconnected beamlines, and knowledge production (Bowker & Star, 2000). The beamline infrastructure, coupled with the multi-aggregate relationship between the scientific instruments and the user community, is key in understanding the facility. It is crucial to the research process, shaping collaboration, how experiments are conducted, data are collected, and knowledge is produced. However, it is often rendered invisible in the public mind and in large-scale quantitative analyses perhaps due to challenges in overlooking the physical research infrastructure and the difficulty of identifying relevant metadata from standard publication aggregators and providers. This perspective allows for the exploration of new perspectives within, but perhaps not exclusively, quantitative studies of science publications of user-oriented Big Science.

# 4. Method

## Data

The ESRF was selected as a case due to its characteristics as a leading research facility in Europe and the world, with thousands of users every year, resulting in an extensive pool of publications with thorough metadata collection and availability. A decision was made not to include more facilities for a variety of reasons. First, the ESRF-ILL Joint Library provides a wealth of data, with over 30,000 publications from 1984 onwards (EPN-Campus, 2014), which provided a rich testbed where the approach of the thesis could be explored and developed. Second, comparing facilities can be challenging because of differing observation sizes and the focus on beamlines, which might vary across facilities. Third and final, the analysis targets not facility itself but its beamlines. The number of beamlines the ESRF hosts is unparalleled across Europe, bringing forth a sample space with a wide variety of cases.

The data collection was done through the ESRF library, which contains all the publications and material related to the facility, as well as the neighbouring institute ILL (EPN-Campus, 2014). It contains a collection of documents, including reports and scientific publications related to the facility. Additionally, the ESRF library augments its publication data, offering insights into various facets related to facility usage. These include which beamline was used for the study, whether internal scientists or facilities were involved in the publication, and whether another beamline or facility has been used in conjunction for the study.

Using the ESRF library for data collection addresses a major challenge encountered when gathering data from similar large-scale facilities catering to external users. As most of the users are not affiliated to the facility, relying solely on affiliation data would exclude a major part of the scientific output of the facility, and bias the analysis toward the beamlines where external users collaborate heavily with internal scientists. A scientist without formal ties, such as co-authorship, to someone affiliated with the facility would see their work omitted from the analysis. In other words, this link is not formally addressed in the publication metadata, so it is invisible to bibliometric data aggregators. There are internal attempts to link external authors with the facility. Most facilities, including the ESRF, now ask for a mention in the acknowledgements section. This often takes form of a sentence or two acknowledging the use of a certain beamline at the facility, sometimes mentioning internal scientists by name. However, it is still not possible to collect all the publications with this method from data services like Web of Science, Scopus, or Google Scholar.

To address this issue of data collection, the DOI of each publication obtained via the facility library is matched with data from the Web of Science to make use of their systematised and structured data, common in bibliometric studies. In the case of this study, relevant variables were kept from the two databases to enrich the analyses of the four articles[2]. Data were collected in two batches. The first collection was done in October 2019, and represents the bulk of the over 35,000 publications of the ESRF. The second batch was collected in 2022 and served to keep the database up to date. This was necessary for Articles III and IV, which needed richer metadata. For Article III, this meant detailed author–affiliation links for the 2011–2021 period. For Article IV, the year 2019 was used for the analysis, as it provided more data in acknowledgements sections, the main topic of study.

In all, the combination of a highly structured and cleaned database from Web of Science, with additional metadata from the ESRF library (EPN-Campus, 2014) provided a rich dataset for analysis with additional metadata, such as beamline names, the involvement of in-house scientists,

2 The ESRF is not the only facility with their own data collection practices. Others include other synchrotron facilities like Diamond Light Source in the United Kingdom and MAX IV in Sweden. This approach can be applied to these and other research facilities if they keep track of their DOIs and other data relevant for the analysis.

and whether other beamlines or facilities were used. This opened several potential avenues for exploration and analysis. Together with additional metadata from Web of Science, this would provide information about disciplines, co-authorships, author affiliations, and the acknowledgements section in text form. Together with the computational methods developed and applied in this thesis, a multi-layered analysis at scale is possible. Some methods have been made available as open-source Python scripts so that others can replicate the analysis, apply them to their own research problem, or improve them[3].

## Empirical approach

The empirical approach attempts to combine the perspectives of metadata and documents from Information Studies with the rigorous and reproducible characteristics of quantitative approaches, accentuating the benefits of both approaches. In general, the methodology adopts an approach where information about user communities is systematically extracted and analysed from journal publications. This is influenced by the thesis placement within bibliometrics, and scientometrics as well as tools from data science where statistical and computational techniques are woven into a larger framework within the Big Science domain, addressing specific questions about the facility (Blei & Smyth, 2017; Desai et al., 2022; Donoho, 2017).

The empirical approach infuses a data science and bibliometrics approach with relevant theories to enrich the analysis of user communities (explored in *Placement of the thesis)*. One goal is to gain a better understanding of beamlines, their configurations, and their role in the user communities within the facilities. Another is to improve the understanding of processes for research with synchrotron radiation, and the potential of this approach to be applied in other areas. The articles employ different methods to explore the facility, which range from statistical analyses common in bibliometrics toward computational and state-of-the-art AI-driven methods. The methods used in the thesis reflect the process of learning

3 See the individual articles and https://github.com/soderstromkr/ for more information and the available scripts.

through the project, and the knowledge gained in both the theoretical and technical understanding of the field.

All the articles in the thesis use or apply statistical and algorithmic methods to a lesser or greater degree. However, *Generic instruments in a synchrotron radiation facility* (Article I) is the more traditionally statistical project, in which an adapted Simpson Index and Herfindahl-Hirsch Index was used to explore multidisciplinary levels of the beamlines in the ESRF for the period 1996-2018. In this article, the concept of generic instruments is operationalised as the disciplinary diversity calculated by the index.

Article I explores the importance of analysing the individual instruments at the facility and attempts to understand their generic (Joerges & Shinn, 2001; Rosenberg, 1992) quality in a quantitative way, focusing on their use by different scientists in different disciplines. It launched the perspective of instruments in user-oriented Big Science facilities for the rest of the project, finding that instruments are used by different disciplines and the difference is relevant for further research. The rest of the articles used statistical and algorithmic approaches as ways to explore and/or validate the different applied methods such as network analysis, geographical mapping, and Large Language Models (LLMs).

In *The structure and dynamics of instrument collaboration networks (Article II),* statistical analyses are used to evaluate the structure of and the changes to instrument collaboration networks within the facility, a concept defined within the article, by calculating network statistics over the period of analysis and measuring how they change over time. In *Global reach, regional strength: Spatial patterns of a big science facility* (Article III), the focus lies on mapping author affiliations for teams of beamline users, and statistical methods are used to analyse distances between geographical coordinates to understand the geographical dynamics of the facility. Finally, in *Who did what? Creating structured data from acknowledgements text with large language models* (Article IV), the focus is on extracting structured data from text. Precision, recall and F1 metrics—often used to evaluate the performance of classification models— are used to validate the use of LLMs as a method, as well as to explore the newly created data.

In parallel to work in Article I, a preliminary exploration of co-authorship networks showed difference between some beamlines, some to a greater degree than others. After some further exploration, it seemed—at least from visual inspections—that the main difference between the network structures revolved around the presence of in-house scientists in the networks. Social Network Analysis (SNA) was the main method for *The structure and dynamics of instrument collaboration networks.* In the article, the concept of instrument collaboration networks is problematised as the formal collaboration between scientists, as made explicit by groups of co-authors that could be grouped around the same beamline. The central idea and motivation for this method was that beamlines could have certain characteristics that result in diverse types of collaboration networks based on theoretical concepts that explore knowledge sharing within social networks. This was visualised and analysed with scripts developed in the programming language Python using existing libraries to calculate network statistics and visualisations. The global and local network properties are calculated for the 2000–2018 period, providing insights into how the structure changes over time.

The methodological approach for *Global reach, regional strength: Spatial patterns of a big science facility* was influenced by a common issue encountered in spatial studies: the need to fit data into a fixed level of aggregation. An opportunity to tackle this problem was presented with WoS data, which included clear author affiliation links since the late 2000s. The method included developing an algorithm to clean and parse the affiliations through a geocoding API (Application programming interface). It was beneficial in two ways. For one, it allowed for the possibility of mapping locations in a Cartesian plane, where measurements of distance could be calculated, as well as translated into interactive geographical maps for the period 2011-2021. The second advantage came in the form of data cleaning, solving a common issue of address capturing: two people are likely to refer to the same place differently in the address field, which the algorithm solved as the API translated the natural text into known places and coordinates. The complete methodology includes additional data handling and analysis including address splitting, geocoding, and calculation of distance measures and mapping, which are explored with more detail in Article III of the thesis. This method is also the first one to be made available to the public as a Python package[4].

The use of machine and deep learning methods to extract information from text data was tested at the beginning of the project. These initial attempts were performed in 2019–2020, using models like GPT and GPT-J (an open-source variant), and not successful, the idea was left in the background for a couple of years. However, the public releases of GTP-3 and eventually ChatGPT reignited interest in the idea. The new attempts showed remarkable progress from just two years earlier, as the methodologies were easier and faster to perform. The method was also inspired by results from previous articles, which analysed formalised collaboration (co-authorship). If beamline scientists were not mentioned as co-authors, there was a chance they were mentioned in the acknowledgements.

This was explored in *Who did what? Creating structured data from acknowledgement texts with large language models.* After initial data inspection it was clear that many other aspects of the research process were also mentioned in the acknowledgements, such as funding, tools and techniques used, beamlines, and facilities. However, this is undoubtedly a more recent phenomenon, as newer publications often contain richer acknowledgements text. LLMs have no established methodological approaches for exploring unstructured text like the acknowledgements section. One exception is a paper about few-shot learning by T. B. Brown et al (2020), where a type of simplified machine learning fine-tuning inspired one of the two methods in Article IV. The second method was influenced by how the public mostly interacts with ChatGPT and other LLMs, which is simply by giving it instructions in natural language. This method was used to extract information from the acknowledgements in 2019 publications, into categories relevant to the analysis. This included the following categories: funding agencies, individuals, physical infrastructure, beamlines, grant numbers, and type of assistance, based on previous related work (Smirnova & Mayr, 2023).

4 https://soderstromkr.github.io/geoaddress/

# 5. Summary of articles

This section contains the summaries of the articles attached in this thesis. As summaries, they provide an overview of the approach and the results obtained with each article, dealing with various aspects of metadata available in the scientific documents in the ESRF database. Each article ends with a short exploration of how the article results are reframed within the context of the theories employed in the thesis. However, it is in the *Discussion* where these ideas are further explored and organised into the different themes found within this new framing.

## Article I. Generic instruments in a synchrotron radiation facility

The article is an exploration of the genericity of the beamlines at ESRF, using bibliometric data to categorise beamlines based on how widely they are employed across different disciplines. It is the first attempt to examine beamlines as a unit of analysis instead of the facility and kicks off this change of perspective for the thesis. The framework of this article originally conceptualises beamlines as generic instruments (Joerges & Shinn, 2001; Rosenberg, 1992), which contain technology originally designed and created for a single, or specific, purpose but since then has broadened its use and application into new fields and sectors. The concept is operationalised as the multidisciplinary level of the beamlines with a diversity index (Herfindahl, 1950; Rhoades, 1993; Simpson, 1949). The objective is to categorise beamlines according to the extent of their use within and across different disciplines. This is argued to be central for a better understanding of how synchrotron radiation facilities are integrated into scientific user communities and how they are used.

The aim of the article is to make sense of the quality of genericity, how it varies between instruments, and how this can be understood in terms of publication metadata related to research fields and disciplines. It uses data from the ESRF publication database, enriched by data from Web of Science. The sample size is a total of 11,218 journal publications for the 1996–2018 period. The combined dataset includes the beamline name, available from the ESRF library database, which makes an instrument-level analysis possible. The methodology compares two approaches: The first calculates the index by journal names; the second provided a top-down classification structure by subject categories in Web of Science.

The results show that generic instruments can be identified, at least in part, by these properties of diversity and concentration of use across different disciplines. Thus, an instrument is generic if the body of research associated with that instrument is diverse and not concentrated in a small range of disciplines. The article provides quantitative evidence that beamlines are generic instruments; that the differences in use between the instruments within the facility are not trivial; and that analysis on the instrument level is fruitful when performing research on synchrotron radiation facilities. The analysis shows examples of which disciplines or topics are related to user communities using these instruments.

The findings also provide some initial insight into the central premise of the thesis. It highlights the richness of data and information when analysing facilities like the ESRF on this granular level. This stands in contrast to the more traditional view of scholarly collaboration in which the types of Big Science facility in focus are mostly high-energy physics intensive, and/or in which the whole facility is taken as the unit of analysis. In this thesis, the categorisation of beamlines is based on the extent of their use across the disciplines using diversity indices. It is a first, but crucial step in understanding the relationship between the configurations of the beamline and the user community, as it illustrates how beamlines in fact attract different user communities and combinations of disciplines. This first step does not investigate the differences of these configurations, what they are or how they relate to the user communities. This is accomplished in the explorations of the social, geographical, and process perspectives explored in the following research articles.

## Article II. The structure and dynamics of instrument collaboration networks

After revealing the various levels of genericity between beamlines, another question seemed relevant to pursue: If the disciplines vary between the beamlines and their multidisciplinary levels vary as well, does the structure of collaboration vary, and does it vary between beamlines? Why do collaboration structures vary? What are the potential causes and effects of this variation? The structure and dynamics of instrument collaboration networks was written to explore those questions. Specifically, the article examines collaboration networks between different beamlines. The focus is on how much the user community is formally collaborating with the in-house scientists while using these beamlines for their research. Specifically, it analyses the relationship between the level of internal (in-house) and external use (user community) of beamlines and how this relates to the structure of their collaboration networks.

The study also explores the formation and change of the networks between different scientific instruments and investigates some reasons behind the different network structures. The dataset for the article consists of co-authorship data for 8,323 journal publications over the 2000–2018 period that used ESRF beamlines in their research. Publication metadata—specifically co-authorship data—is used to construct the networks surrounding each beamline in the analysis. Some statistics are calculated over the period to determine the structure of their networks, and how they change including the number of nodes, the number of edges, the average degree, the number of components, and the giant component size, which serve as a proxy for structure. Co-authorship serves as a proxy for formal collaboration, where authors will be connected if they appear in the same co-authorship list. As the data includes 31 beamlines, three beamlines (ID17, ID19 and ID23-1) that show significantly different network structures are selected for further analysis and serve as a baseline for further contextualisation and examination, inspired by the findings Article I. The concept of instrument collaboration networks is introduced, where scientists positioned in central parts of the network should contain knowledge and know-how from the different teams of scientists in their proximity

(Burt, 2004; Collins, 2012; Granovetter, 1973; Polanyi, 1962), pertinent to specific technological characteristics and related practices (Shrum et al., 2007).

The results of the study suggest that there is a significant relationship between a high percentage of 'external-only' user teams and a significant level of fragmentation in the network structure. This results in a disconnected network with only 'within-team' connectivity (connections with their immediate team members, but not to other teams). Cases where instruments have a lower ratio of external-only user teams are associated with more interconnected collaboration networks, which show high levels of connectivity between teams. The analysis of the three beamlines suggests that overall, technological improvements that increase ease-of-use for user communities might affect the formalisation of co-authorships with in-house scientists, providing insights into the factors that shape collaboration networks, and the ways that different teams interact and share knowledge.

Automation and remote access capabilities were suspected to be main factors in these dynamics but were not fully explored beyond the text descriptions presented in the article that highlighted some of these properties. Furthermore, the article provides insights into which conditions are necessary to formally recognise co-authorship in scientific work. The results add additional context to the thesis, as they provide a first look into empirical quantitative evidence, suggesting that the technological configuration of the beamline could influence how they are used. With the conceptualisation in the article about the effects of technology in science and the beamline in mind, technology shapes not only user access in terms of disciplinary use, but also in terms of collaboration patterns. That is, the willingness to formalise collaborative ties with beamline scientists in terms of perceived work by the external team further shapes the user community.

## Article III. Global reach, regional strength: Spatial patterns of a big science facility

So far, two dimensions of beamline characteristics—the disciplinary dimension— provide information about who uses the beamline, which in turn says something about its genericity or versatility; as well as the social dimension, which offers information about the collaborative structure between the different beamlines. One final dimension to explore is that of geography, as hints of remote usage were found in the results of the previous article. New questions were beginning to form, such as: Where are the users coming from? How are their teams arranged over geographical space and time? This article presents a methodology for extracting and analysing author–affiliation data and attempts to provide answers to these questions.

The aim of the article is to provide a methodology to understand the geographical reach and spatial patterns associated with publications from ESRF. The data consist of 17,870 journal publications over the 2011–2021 period, resulting in 76,850 total affiliations of which 11,120 are unique locations. The methodology is systematic and algorithmic. Using the author affiliation addresses from publications, it extracts, cleans, and geocodes them to generate geographic coordinates. Addresses are disaggregated from author names, then geocoded via Google Maps API. These coordinates are used to construct dynamic maps and calculate spatial statistics; namely geographical distances between author teams and between teams and the ESRF. The article uses publication data from 2011–2021. As a mostly methodological contribution, the scripts and resulting analysis including detailed, interactive spatial maps at multiple scales, and more dynamics are available as an open-source python package .

Overall, the results show the growing international scope but also regional concentration of ESRF author affiliations with presence across Europe, the Americas, Asia, Africa, and Australia. There is a strong concentration of affiliations in Western Europe, especially around Grenoble, Paris, and London. Geographical distance increases over time, and some noticeable differences are seen between instruments in their spatial patterns and evolution. Differences in spatial patterns are observed across beamlines - some have strong regional agglomeration while others are more dispersed. The results show how mapping author affiliation using the proposed method can be used to understand the geographical reach and dynamics of the facility and its beamlines. It observes differences in spatial patterns across the beamlines, however, with a general increase in the reach of the facility. The results from this article further show that the configuration of the beamlines shapes user communities, and with this evidence, it shapes the user community geographically. During the latter part of the period of analysis, more remote access was reported in the ESRF Highlights for 2022 (2023), which catalysed during the COVID pandemic, and appears to be reflected in the results in the article.

## Article IV. Who did what? Creating structured data from acknowledgement texts with large language models

The last article in the thesis revisits the idea of formal collaboration and explores the following question: If certain technological characteristics decrease incentives for users to acknowledge collaboration formally, can we find that information elsewhere? This article leverages LLMs to extract structured data on funding, collaborators, facilities etc. from unstructured acknowledgement texts of publications related to the ESRF. The background of the article revolves around how evaluation and the analysis of science often focuses on measuring output, like citations and publications, which provides limited insight into the processes behind science. The acknowledgements section provides a unique window of possibilities into these processes, but free-form text is difficult to analyse at scale. Thus, it explores the use of generative AI, specifically LLMs, as tools to structure data from the acknowledgements section of scientific publications.

The aim of the article is to uncover the potential of LLMs for quantitative analysis of research processes and their implications for science evaluation and policy. The article also reveals some of the informal processes that led to the scientific outputs of the ESRF. Text data from 1,482 journal publications in 2019 was extracted from acknowledgements sections, where the year 2019 was chosen, as newer metadata contained acknowledgements that are more detailed. The data includes the DOIS and the Funding Text, which captures the acknowledgements provided by the authors. The methodology of the study involves using LLMs to structure data using two approaches, an "instruction" approach and a Few-Shot Learning (FSL) approach. The term "instruction" refers to writing free-form text instructions, prompting the LLM to retrieve specific information from the acknowledgements text, a form of prompt engineering. It provides specific queries so that the model extracts data related to funders, individuals, physical infrastructure, beamlines, grant numbers, and types of assistance. On the other hand, FSL involves training the LLM with a few examples of prompts and completions to contextualise the required results. It provides mappings between example inputs and outputs for each category from which to extract information. The model is then applied to each of the acknowledgement texts to retrieve the relevant information. The results of the study are evaluated based on precision and recall measurements for each category.

The results show a wide range of funding agencies, beamlines, people, and types of assistance provided to the authors of the publications, uncovering the range of entities supporting the research process 'behind the scenes.' Often, the people mentioned in the texts are ESRF in-house scientists providing some sort of support. For instance, they set up the experiment or grant access to the beamline. Other types of assistance included financial, technical, conversational (discussions) support as well as beamline access. While limited in terms of observation size, these results provide preliminary quantitative insights into the diverse processes at play in the collaborative ecosystem of the ESRF. The results highlight the individuals, technology, and assistance that are not visible in current metadata from services like Web of Science, which this thesis uses as one of the two main data sources. It also illustrates the diverse processes and informal collaborations that take place at the beamlines and facility, both within the ESRF and including other beamlines and individuals in other facilities, like DESY (Deutsches Elektronen-Synchrotron) in Germany. The information extracted from the acknowledgements text highlights parts of the research process that risk remaining hidden from traditional metadata and could be used as a method to uncover the dynamics of the facility that remain hidden in other metadata.

# 6. General discussion

Overall, the results reinforce the perspective that beamlines have different technological configurations that result in different user communities (Article I). Some beamlines show publications where the in-house scientists are formally involved, such as the physics-oriented beamlines in the realms of hadrons and nuclei (Article II). It also shows some beamlines, within structural biology for instance, which cater to large user communities external to the facility and do not formally involve in-house scientists or technicians through co-authorship (Article II). However, others are hosted by scientific fields and disciplines that have, or aim to implement, instrumentation that enables them to be used by others. These scientific instruments are initially used for a specific field of science or study. However, some can be expanded to other fields, and spaces between fields (Article II).

The two following discussion points address the research questions, drawing on the results and the framing to highlight beamlines as unique collaborative spaces for research, and a discussion that explores the mechanisms through which their technological configurations attract different combinations of user communities. After this, the discussion section broadens to some implication on the aggregate facility, and other user-oriented Big Science, with a discussion on the visibility of research infrastructures. A discussion around metadata follows. Finally, some limitations are addressed.

## Beamline spaces

Article I pivoted the analysis of the ESRF to the granular perspective of the beamline, as it is the place where most, if not all, user experiments take place. However, contrary to the initial portrayal in Article I, beamlines are not just isolated instruments. This seems evident if we recollect how these facilities function, as explored in the Introduction section. At the ESRF, the energy utilised by all beamlines originates from the facility's larger structural components. Electrons, produced and accelerated in the linear accelerator and booster ring, are transported to, and stored in, the storage ring. They are then directed and transformed into different beamlines with unique technological configurations. Beamlines serve user communities with different disciplines and combinations of disciplines, as shown by the index calculations in Article I, setting the stage for the subsequent articles, further examining the social and spatial characteristics—as well as the processes—that are created and enabled by these spaces.

Drawing inspiration from studies on Information Studies, STS, and studies of Big Science, the revised framework of the thesis proposes that a beamline is more than a scientific tool shaping knowledge. Instead, it seems more appropriate to analyse the beamline as a space, as detailed in the Theory section and evidenced by the findings in the original research articles. As such, the beamline space provides a set of scientific instruments (Price, 1984) a common space, language, and set of tools for scientists from different disciplines (Bowker & Star, 2000). They share robust identity across sites (Bowker & Star, 2000; Star & Griesemer, 1989), provided physically by the large storage ring, and by the expert knowledge of the in-house scientists (Article II). They can adapt to local needs and constraints, as different spaces can be used to study a variety of disciplines (Joerges & Shinn, 2001; Rosenberg, 1992), from biological structures and materials sciences to nuclear and hadron physics (Article I). The collaboration enabled in these spaces helps build relationships between scientists and has the potential to break down disciplinary boundaries (Article II).

With the barriers of discipline (Article I), geography (Article III), and language (Galison, 2010) being continually (Arthur, 1990) eroded, scientists worldwide can make use of the capabilities of these beamlines. This global engagement is further reinforced by technological advancements (Arthur, 2009) and the dedicated efforts of beamline scientists. The dynamic nature of the beamlines is from their adaptability, driven by the evolution of scientific instruments within them. This adaptability not only caters to current research needs but also holds the promise of attracting newer user communities, ensuring the sustained relevance of the facility in a competitive scientific landscape (Articles II and IV).

This perspective of the beamline aids in recognising their heterogeneity within the facility as essential, and that treating all beamlines as equals overlooks their specialised roles in user communities. For now, it has only been hinted that beamlines have different combinations of scientific instruments, and their different configurations will shape user communities as they parse users from differing disciplines and practices, which then converge and form the existing and potential user communities. However, user communities also shape these technological configurations because they modify, change, or improve according to their current and emerging needs. This appears to be a mechanism in which the beamline scientists and user communities negotiate, directly or indirectly, the configuration of the beamline spaces to remain relevant to their user community (D'Ippolito & Rüling, 2020; Hallonsten & Heinze, 2012). The way beamlines are configured and the relationship to the user community are discussed further in the following section.

## Technological configurations

One crucial element of research with synchrotron radiation is brightness. To put it simply, a brighter beam enables scientists to explore more details. Thus, an extensive amount of work and funding goes to improving the capabilities of the storage ring to increase the brightness and hence, also the potential scientific performance and output of the facility. The latest example of these efforts is the construction of the Extremely Brilliant Source (EBS) during the 2015–2020 period, at a cost of 150 million Euro, which improves the potential energy output for all beamlines (Chenevier et al., 2021; Raimondi, 2016), but also means upgrading the beamlines to accommodate this newfound power.

Another key aspect for some beamlines is being able to manage a variable energy range. A tuneable beamline is one that can adjust the incoming beam from the storage ring to a wider range of energy levels, which means that beamlines with a tuneable range can accommodate a wider range of experiments. However, the technological configuration of the beamline goes beyond just energy levels and ranges (Weitkamp et al., 2010). That is, instead of instruments that directly contribute to the energy intensity or range, they contribute to the ways users interact with the beamline.

Two main additional aspects are identified—automation and remote access—achieved via software and hardware upgrades in the instrumentation of the beamlines. These upgrades and combinations of technology have the potential to increase the accessibility of these beamlines (Article I) within and beyond existing user communities (Article II) to more users, including users with less experience with this type of technology and science (Chenevier et al., 2021; ESRF Highlights 2022, 2023; Flot et al., 2010; Gabadinho et al., 2008). Beamlines contain instruments that allow some, or all, processes to be automated, as well as instruments that allow some processes to be done remotely, and some via mail-in sampling, without the need for the user community to step into the laboratory setting. This has the potential to lead to new user communities, which can be rewarding for both the facility and the user community, as they bring in new knowledge.

The combinatorial (Arthur, 2009) addition of instruments has the potential to increase its ease of use. This includes improvements like energy output and range, as well as automation and the implementation and adoption of remote use, as explored in The facility and its beamlines. For instance, ID19, one of the oldest and longest running beamlines at the ESRF, with a history of constant energy range and automation upgrades, shows fragmentation of its formal collaboration network over time (Article II). However, it remains one of the most used beamlines in the facility (European Synchrotron Radiation Facility, n.d.i; Weitkamp et al., 2010), implying changes in collaboration between the user community and in-house scientists (D'Ippolito & Rüling, 2019).

Expectations between beamlines should be balanced appropriately, since not all may have the capacity to increase their user communities in the same way. For instance, Article II shows that the medical beamline, ID17, used for diagnosis and irradiation therapy research requires in-situ experimentation. As the beamline hosts a consistent and homogeneous user community, it should not be expected to grow in the same way as the multidisciplinary beamlines like ID19. In beamlines like ID17, one can expect formalised collaborative practices (D'Ippolito & Rüling, 2019).

The dynamics between beamline configurations and user communities are initially examined in Article II, where a selection of beamlines was shown to include different technological characteristics, including hardware and software, which influence how the beamline is used and by whom. Some beamlines are shown to be highly versatile over the period, which influence the collaborative structure of the beamline and the possibility to reach the geographically distant user communities exemplified by Article III.

The mechanisms through which formal collaboration is shaped occur via various methods of interaction between the user community, mostly external to the facility, and the in-house scientists, as well as the perception of concrete working practices due to automation and remote capacities (Article II). This offers quantitative insights into tensions between formalising collaborative practices and expertise with beamline scientists via the tool of co-authorship. This might include the difficulties of building user friendly interfaces for the user community (Gabadinho et al., 2008) and the perception of work of the in-house scientists by the user communities (D'Ippolito & Rüling, 2019), which seem to change how collaboration is formally, or informally credited as shown in Articles II and IV. Refining the method in Article IV could reveal the extent to which mentions of remote work have increased in the acknowledgements. ESRF, mirroring reporting noting an increase in remote use by beamlines in the structural biology group (ESRF Highlights 2022, 2023).

There are consequences to digitalisation to consider or manage. For instance, while more users, from different communities join the research community at the facility, less formal collaboration seems apparent, which could be due to the reduced personal interaction with beamline scientists, and/or different collaboration practices of the newer disciplines (D'Ippolito & Rüling, 2019), also shown in Article II. Article IV shows the extent of informal means of crediting work by other scientists, as well as the beamlines and other facilities, which is often not one of the main visible metrics when analysing and evaluating collaboration or contributions.

The efforts to improve these different technological and social aspects of the beamline happen at different timescales. While there is indeed a constant attempt to increase energy ranges at the facility, automation and remote use could be described as more recent and emerging efforts. The former sees increased interest during the 2000s, and the latter a becoming much more established during the later 2010s but accelerated during the COVID-19 pandemic (Chenevier et al., 2021; ESRF Highlights 2022, 2023). These technical additions to the facility seem to follow a progression of beamline improvements, brought upon by constant internal and external factors, and a means of adaptation (D'Ippolito & Rüling, 2020; Hallonsten & Heinze, 2012; Westfall, 2008). As part of a wider ecosystem, facilities and projects adapt to external and internal forces. They compete with others for funding and the attention of the scientific community, as well as the imagination of the public. These adaptive capabilities have implications for the study, analysis and/or evaluation of the ESRF and other user-oriented Big Science, explored in the next section.

## Infrastructural adaptation and visibility

Big Science facilities have historically exhibited a remarkable capacity for adaptation and development that enables their continued usefulness and survival over extended timeframes, encompassing more than just technical upgrades to the instrumentation presented in this thesis (D'Ippolito & Rüling, 2020; Hallonsten & Heinze, 2012; Heinze & Hallonsten, 2017; Westfall, 2008). For instance, synchrotron radiation facilities have progressively expanded their capabilities far beyond what the pioneering developers envisioned in the 1990s (Hallonsten, 2016a). Similar to the ESRF, DESY facility in Germany provides another illustration of transformational adaptation from single-mission particle physics towards a multi-mission research centre including photon science (Heinze et al., 2015, 2017).

Facilities also adapt to geographical and organisational dimensions, as evidenced by the strategic partnerships between the ESRF and the neutron research centre ILL in Grenoble; and between MAX IV laboratory and the European Spallation Source (ESS), and Lund University (Hallonsten, 2012; Hallonsten & Cramer, 2020; Rekers & Sandell, 2016). Some of these determined collaborations are visible in Article III, where some of the user communities are anchored to the region surrounding the ESRF, and the mentions of other beamlines and facilities in Article IV. Technical advancements to the beamline infrastructure in one facility may diffuse to others as well. Furthermore, upgrades implemented at one facility propagate benefits to peer organisations operating related infrastructure, like the enhancements to neutron-scattering instruments at ILL which benefit the broader neutron facility landscape (D'Ippolito & Rüling, 2020).

The inherent adaptability of these research infrastructures differentiates contemporary facilities from the large-scale mission-driven high energy physics machines, which often fell victim to technical, political, and/or organisational obsolescence upon achieving their objectives (Hallonsten, 2016a). However, the objective of perpetual adaptability of today's facilities (D'Ippolito & Rüling, 2020) may also face limitations. Derek de Solla Price and Alvin Weinberg presciently warned that endless exponential growth could jeopardise scientific serendipity and academic autonomy (Price, 1963; Weinberg, 1961). Yet, for the present moment, the open-ended versatility found in the results in this thesis sustains the ESRF as a leading synchrotron radiation facility, extending the value of user-oriented Big Science, as they creatively expand capabilities and specialisations to serve user needs. Their responsive and proactive evolution is proving crucial for their continued existence—and integration—with the broader scientific and innovation ecosystems. This adaptability could be achieved via the distributed heterogeneity of the beamlines, which allow the facility to gather varied information about their diverse user communities via the beamline, which they can then use to develop the necessary instruments to remain relevant (Benbya et al., 2020).

This adaptability remains partly unseen in quantitative research policy and evaluation that relies on traditional metadata sources. This results in making the infrastructure invisible, although it is this infrastructure that partly enables the scientific work in the facilities. This omission perpetuates the criticism surrounding how these facilities are funded, built, and then left to their own devices to survive (Cramer et al., 2020; Hallonsten, 2016a, 2020b). Furthermore, the dynamic, customised beamline infrastructure within the facility makes it difficult to identify who the users are. However, the current framing has the possibility to uncover the relationships observed between technology and user communities, partly invisible to the current analyses and evaluation. Article IV shows that some dynamics are only partly visible, since they can be found in the acknowledgements section, an emerging field in bibliometrics. Article IV also shows these acknowledgement practices are relatively new, so past interactions could be effectively unaccredited in scientific publications, unless they could be found within the full text. However, they remain unaccounted for in large metadata aggregators, as they do not make up the core of variables commonly included in bibliometric or other large quantitative analyses.

The methods in the thesis could serve as a basis to study other complex, customised research infrastructure ecosystems. The ESRF is indeed a global facility with a large user community and state-of-the-art technology—but is not the only one. Other user-oriented Big Science facilities or projects that exhibit similar dynamics abound, where an external user community interacts with internal scientists and technicians in a space akin to the user laboratory. Although they might differ in terms of technology, discipline, and practices, other facilities and projects also serve unique conditions for research for their users. Other synchrotron radiation facilities are the most obvious candidates, and indeed, they contain beamlines that serve user communities. Other user-oriented Big Science facilities include free-electron lasers, neutron sources, high-magnetic field laboratories, data centres, and telescopes. Even more novel space industry efforts could apply as well. For instance, the European Space Agency (ESA) has collaborated with the Spanish synchrotron radiation source ALBA to develop a beamline that recreates optical conditions in space to support the development of a telescope (Heinis et al., 2021). Space agencies, private and public, have also deployed experiments in various disciplines like materials science, biology, and physics that benefit from low gravity environments (Virgin Galactic, n.d.; Webb, 2020).

## On metadata

The primary contribution of the thesis to LIS is through the exploration of metadata and methods in the individual research articles that extract and analyse several sources of metadata, both structured and unstructured, enabling the effective organisation of information (Alemu & Stevens, 2015a) from the facility. This includes the beamline name captured by the ESRF-ILL library, and the matching made possible by the collection of DOIs by both the library and Web of Science. For the articles, metadata included the Journal Names and WoS Subject Categories for Article I, the co-author lists for Article II, the author affiliations for Article III, and finally the acknowledgements text for Article IV. This has enabled the creation of a robust picture of the beamline, and the ESRF, where fragments of coded and full text metadata have explored the people, technology, and processes that make science possible.

The thesis aligns itself with an LIS perspective that favours flexibility (Åström, 2006; Petras, 2023) surrounding the nature and analysis of information, documents, texts, data, and the techniques used to capture, represent, organise, and retrieve them (Bawden & Robinson, 2012b; Buckland, 2012). The informational framing of the beamlines provides a distinct vantage point rooted in Information Studies in which metadata enables the exploration of disciplinary (Article I), collaborative (Article II), spatial (Article III), and informal processes (Article IV) surrounding the beamlines. Specifically, it addresses the value of analysing documents and texts as representations of a process or knowledge, rather than relational or evaluative outputs. The thesis also contributes to a shift in quantitative studies of science publications infusing theory and computational methods to reveal real world processes (Heinze & Jappe, 2020; Leydesdorff et al., 2020; Vertesi, 2019; Wyatt et al., 2015). In the case of the ESRF, it means revealing how beamlines configure user access, practices, and collaborations. Furthermore, the in-house scientists produce additional documentation that can be explored and contribute to the analysis in future work.

All articles in the thesis show the value of embracing the messy granularity of metadata. However, the future of metadata, and the way it is collected, analysed, and interpreted has the potential to undergo a transformation with sophisticated computational methods, including the implementation of LLMs. As explored in Article IV, successfully extracting data from acknowledgements text can provide insights into information that accessible metadata simply does not yet capture, and can enrich existing, but messy, metadata (Alemu & Stevens, 2015b). The acknowledgements text seems to exist in an intermediate zone that offers concise information about the processes related to the scientific publication, at least in the case of the later ESRF publications explored in Article IV, uncovering details not visible based on formal publication or affiliation data alone. Current approaches and solutions to metadata favour sufficiency and necessity (Alemu & Stevens, 2015b), striking a balance between readability, conciseness, and information content. However, being able to analyse a full text—or, as in the case of Article IV—an acknowledgements section, to extract tailored information to enrich the available metadata (Alemu & Stevens, 2015b) has the potential to challenge current metadata services and databases like Web of Science, Scopus, and their analytics services, which provide top-down classifications and aggregations. As with the ESRF, the heterogeneity of science, and that of most domains of inquiry and information in general, resists this type of classification.

## Limitations

There are limitations to keep in mind. Some are mentioned in the individual articles, but some have come up during seminar sessions surrounding the thesis project, and re-readings of the articles considering the new framing and further considerations arising from the thesis writing process. Below, the limitations for each article are discussed.

One challenge persisted across the articles, which was the naming of the different beamlines across the ESRF. As discussed in Article I, beamlines undergo a life cycle, some longer than others. In some cases, older beamlines are upgraded, moved, or replaced by newer beamlines, which sometimes leads to changes to their names. This is reflected in the ESRF Library database when a beamline has a certain suffix. In Article I, a decision was made to try to wrangle some of that heterogeneity into a collection of more stable beamline names, under the assumption that the initial part of the beamline would serve as the base for the overall "family" of beamline suffixes. As the analysis aggregated an extended period (1996-2018), it was deemed necessary to gain some sort of consistency across beamline names. For instance, beamline ID23-1 and ID23-2 would be part of the same beamline area and would be aggregated into beamline ID23. In fact, the technical specifications for the ID23 "Gemini" beamline indicates this is the case (European Synchrotron Radiation Facility, n.d.-e), as it is for beamline ID14, in which the suffixes 1–4 indicate different stations for user groups (European Synchrotron Radiation Facility, n.d.-c) (See Article I for more detail). However, for the subsequent articles that dealt with changes over time over shorter periods a less hands-on approach was selected, where the beamline names would be treated as they are, and no aggregation would be attempted.

The second limitation concerns the lack of the time component in Article I, which could have investigated how the different combinations of disciplines change over time. This was included in Articles II and III. However, Article IV also only considers one year, 2019, for the analysis.

A third limitation for Article I was not mentioned in the article but came up in a seminar discussion. Often, the names of the journal or the subject categories have words or names that imply a multidisciplinary journal, or a multidisciplinary category (e.g., the journal "Nature" or the category "Materials Science"). This means that there is a chance that the index is under-representing the genericity of the beamlines, and in fact could be even more generic than previously thought. A suggestion for future research could be to consider the names that imply multiple disciplines with some form of multiplier. Since it is not possible to know, from a large-scale approach, exactly how many disciplines are included in multidisciplinary journals and categories, the exact multiplier should be derived carefully.

A fourth limitation revolves around the compositions of teams, either socially (Article II) or geographically (Article III). This is the case when there is no collaboration with a scientist affiliated with the facility and only external users appear on the publication as co-authors. In these instances, the publication metadata used for these studies does not indicate who in the team was the point of contact between the user team and the beamline, either by visiting the beamline on-site or by interacting with the remote terminal. Furthermore, there is no systematic way of using publication metadata for telling when the team member visited the facility or interacted with a remote beamline.

However, it can reveals important insights into the composition of teams, their disciplines, how they collaborate, where they work, and who they had contact with at the facility, other facilities, and which beamline or beamlines they used in their experiment. The information gathered from the results can be used to create a social and geographical map of who these user communities are, helping beamline scientists, facility managers, and policymakers with a better understanding who is benefiting from these beamlines and facilities. With more internal data—for instance, the user database from these facilities or the submitted proposals—it would be possible to match who applied and who interacted with the in-house scientists and/or the beamline.

The fifth and final limitation is the novelty of the methodology in Article IV, as it deals with emerging technology and methods in constant development. In fact, it has already shown improvements in the months since submission. While it does serve as an initial attempt to use these models for information extraction in the acknowledgements, LLMs have been upgraded to understand context better, as well as to use functions to interact with other programs and/or scripts. The data could also be cleaned further to make sure repetitions do not skew the results. Furthermore, this data could be further explored, such as a social network analysis of informal collaboration, technologies, and other facilities.

# 7. Concluding remarks

This thesis embarked with the aim to explore of the role of beamlines within the European Synchrotron Radiation Facility (ESRF). These beamlines, conceptualised as spaces for the user communities and in-house scientists, collectively constitute a beamline infrastructure that is pivotal in driving research and fostering collaboration at the facility. The overarching goal was to understand their profound impact on scientific collaboration and knowledge production at the ESRF. To achieve this, the study analysed various metadata derived from scientific publications, shedding light on how beamlines shape access to and foster collaboration across different scientific user communities.

Using a rich dataset from the ESRF library, encompassing over 30,000 publications spanning the period from 1994 to 2021, this study employed computational and bibliometric techniques to provide diverse insights. These perspectives encompassed the disciplinary, collaborative, geographical, and process dynamics associated with the beamlines.

The main contribution of this thesis lies in its methodological approach, adding to the existing scientometrics toolbox and tailoring it to the analysis of beamlines and their user communities. This approach leverages publication metadata, including but not limited to, Journal Names, Subject Categories, Author Names, Author Affiliations, and Acknowledgements. Furthermore, a secondary contribution in the form of a conceptualisation of beamlines—as spaces intricately configured technologically, with disciplinary, social, and geographical implications—offers a novel framework to analyse the ESRF and other user-oriented Big Science facilities through their own user spaces.

Empirically, this study provides quantitative evidence of the disciplinary, collaborative, spatial, and process dynamics that exist between and across the beamlines at the ESRF and their user communities. The results collectively emphasise that the technological configuration of beamlines attract a diverse collection of user communities, which, in turn, foster both specialised and interdisciplinary research. Moreover, the study highlights the significant influence of the continuous beamline upgrades—whether in terms of energy levels, energy range, automation, or remote access—on collaboration and participation among the user community, positioning beamlines as pivotal cross-disciplinary nodes within a broader infrastructure.

However, it is worth noting that the study is not without its limitations, which pave the way for future research avenues. The analysis, while comprehensive across different metadata, is anchored solely on publications. This means it lacks the richness that internal data from facilities might offer, including user data and proposals. Future research could bridge this gap by integrating additional data sources, painting an even more complete picture of the processes behind the beamlines. More detailed comparative analyses, focusing on specific beamlines and examining factors such as configurations, user demographics, and policies, could further enrich our understanding of the beamline. This could also open an opportunity to consider publications that draw upon more than one beamline, an aspect that remains underexplored in the original research articles presented in this thesis.

There is much more to explore. The methodologies and perspectives employed in this thesis are versatile. They have the potential to be applied in the study of other user-oriented Big Science. Policy makers and managers can apply the methods in this thesis to explore the role their beamlines play in attracting user communities. Furthermore, the exploration of metadata and its role in uncovering empirical processes provides an exciting opportunity to explore other domains.

# References


A. Díaz-Faes, A., & Bordons, M. (2017). Making visible the invisible through the analysis of acknowledgements in the humanities. *Aslib Journal of Information Management*, *69*(5), 576–590. https://doi.org/10.1108/AJIM-01-2017-0008

Aarden, E. (2023). Infrastructuring European scientific integration: Heterogeneous meanings of the European biobanking infrastructure BBMRI–ERIC. *Social Studies of Science*, *53*(4), 572–598. https://doi.org/10.1177/03063127231162629

Abramo, G., & D'Angelo, C. A. (2023). What can bibliometrics contribute to understanding research productivity? In *Artificial Intelligence in Science: Challenges, Opportunities and the Future of Research*. OECD Publishing, Paris. https://doi.org/10.1787/2afb7717-en

Alemu, G., & Stevens, B. (2015a). Introduction. In G. Alemu & B. Stevens (Eds.), *An Emergent Theory of Digital Library Metadata* (pp. 1–9). Chandos Publishing. https://doi.org/10.1016/B978-0-08-100385-5.00001-8

Alemu, G., & Stevens, B. (2015b). The principle of metadata enriching. In G. Alemu & B. Stevens (Eds.), *An Emergent Theory of Digital Library Metadata* (pp. 67–69). Chandos Publishing. https://doi.org/10.1016/B978-0-08-100385-5.00005-5

Álvarez-Bornstein, B., & Montesi, M. (2020). Funding acknowledgements in scientific publications: A literature review. *Research Evaluation*, *29*(4), 469–488. https://doi.org/10.1093/reseval/rvaa038

Andersson, C., & Sundin, O. (2023). The elusive search engine: How search engine use is reflected in survey reports. *Journal of the Association for Information Science and Technology*. https://doi.org/10.1002/asi.24819

Arthur, W. B. (1990). Positive Feedbacks in the Economy. *Scientific American*, *262*(2), 92–99.

Arthur, W. B. (2009). *The nature of technology: What it is and how it evolves*. Simon and Schuster.

Åström, F. (2006). *The social and intellectual development of library and information science*. Umeå University.

Åström, F. (2016). *Infrastructures as analytical framework for mapping research evaluation landscapes and practices*. Materiality and social dynamics of information infrastructures and learning. http://lup.lub.lu.se/record/0a186462-ad14-49a3-a54a-9ff972937d61

Bates, M. J. (2005). An introduction to metatheories, theories and models. In K. E. Fisher, S. Erdelez, & L. (E. F.) McKechnie (Eds.), *Theories of information behavior* (pp. 1–24).

Published for the American Society for Information Science and Technology by Information Today.

Bawden, D., & Robinson, L. (Eds.). (2012a). Informetrics. In *Introduction to Information Science: Volume 1: Foundations of the Information Sciences* (Vol. 1, pp. 165–186). Facet. https://doi.org/10.29085/9781783300761.015

Bawden, D., & Robinson, L. (2012b). What is information science? Disciplines and professions. In *Introduction to Information Science* (pp. 1–18). Facet. https://doi.org/10.29085/9781783300761.008

Benbya, H., Nan, N., Tanriverdi, H., & Yoo, Y. (2020). *Complexity and Information Systems Research in the Emerging Digital World* (SSRN Scholarly Paper 3539079). https://papers.ssrn.com/abstract=3539079

Beteva, A., Cipriani, F., Cusack, S., Delageniere, S., Gabadinho, J., Gordon, E. J., Guijarro, M., Hall, D. R., Larsen, S., Launer, L., Lavault, C. B., Leonard, G. A., Mairs, T., McCarthy, A., McCarthy, J., Meyer, J., Mitchell, E., Monaco, S., Nurizzo, D., … McSweeney, S. M. (2006). High-throughput sample handling and data collection at synchrotrons: Embedding the ESRF into the high-throughput gene-to-structure pipeline. *Acta Crystallographica Section D: Biological Crystallography*, *62*(10). https://doi.org/10.1107/S0907444906032859

Blei, D. M., & Smyth, P. (2017). Science and data science. *Proceedings of the National Academy of Sciences*, *114*(33), 8689–8692. https://doi.org/10.1073/pnas.1702076114

Börner, K., Silva, F. N., & Milojević, S. (2021). Visualizing big science projects. *Nature Reviews Physics*, *3*(11). https://doi.org/10.1038/s42254-021-00374-7

Bouvet, T., Deflets, C., Mitchell, E., & Revol, J. (Eds.). (2007). *Science and Technology Programme 2008—2017* (Vol. 1).

Bowker, G. C. (1994). *Science on the Run: Information Management and Industrial Geophysics at Schlumberger, 1920-1940*. MIT Press.

Bowker, G. C., Baker, K., Millerand, F., & Ribes, D. (2010). Toward Information Infrastructure Studies: Ways of Knowing in a Networked Environment. In J. Hunsinger, L. Klastrup, & M. Allen (Eds.), *International Handbook of Internet Research* (pp. 97–117). Springer Netherlands. https://doi.org/10.1007/978-1-4020-9789-8_5

Bowker, G. C., & Star, S. L. (2000). *Sorting Things Out: Classification and Its Consequences*. The MIT Press. https://doi.org/10.7551/mitpress/6352.001.0001

Broadus, R. N. (1987). Early approaches to bibliometrics. *Journal of the American Society for Information Science*, *38*(2), 127–129. https://doi.org/10.1002/(SICI)1097-4571(198703)38:2<127::AID-ASI6>3.0.CO;2-K

Brookes, B. C. (1990). Biblio-, sciento-, infor-metrics?? What are we talking about? In L. Egghe & R. Rousseau (Eds.), *Informetrics 89/90. Selection of Papers Submitted for the Second International Conference on Bibliometrics, Scientometrics and Informetrics.* (pp. 31–43). Elsevier.

Brown, G. E., Jr., Sutton, S. R., & Calas, G. (2006). User Facilities around the World. *Elements*, *2*(1), 9–14. https://doi.org/10.2113/gselements.2.1.9

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., … Amodei, D. (2020). *Language Models are Few-Shot Learners* (arXiv:2005.14165). arXiv. https://doi.org/10.48550/arXiv.2005.14165

Buckland, M. (2012). What kind of science can information science be? *Journal of the American Society for Information Science and Technology*, *63*(1), 1–7. https://doi.org/10.1002/asi.21656

Burt, R. S. (2004). Structural Holes and Good Ideas. *American Journal of Sociology*, *110*(2), 349–399. https://doi.org/10.1086/421787

Callon, M., Courtial, J. P., & Laville, F. (1991). Co-word analysis as a tool for describing the network of interactions between basic and technological research: The case of polymer chemsitry. *Scientometrics*, *22*(1), 155–205. https://doi.org/10.1007/BF02019280

Capshew, J. H., & Rader, K. A. (1992). Big science: Price to the present. *Osiris*, *7*, 3–25.

Chen, C. (2017). Science Mapping: A Systematic Review of the Literature. *Journal of Data and Information Science*, *2*(2). Scopus. https://doi.org/10.1515/jdis-2017-0006

Chenevier, D., McCarthy, J., & Joly, A. (2021). ESRF-EBS: Research and Innovation in the Time of COVID-19. *Synchrotron Radiation News*, *34*(1–3), 18–20. https://doi.org/10.1080/08940886.2021.1902716

Collins, H. (2012). *Tacit and Explicit Knowledge*. University of Chicago Press.

Cramer, K. C. (2017). Lightening Europe: Establishing the European Synchrotron Radiation Facility (ESRF). *History and Technology*, *33*(4), 396–427. https://doi.org/10.1080/07341512.2018.1489762

Cramer, K. C., Hallonsten, O., Bolliger, I., & Griffiths, A. (2020). Big Science and Research Infrastructures in Europe: History and current trends. In O. Hallonsten & K. C. Cramer (Eds.), *Big Science and Research Infrastructures in Europe*. Edward Elgar Publishing. https://doi.org/10.4337/9781839100017.00007

Crease, R. P., & Westfall, C. (2016). The New Big Science. *Physics Today*, *69*(5), 30–36. https://doi.org/10.1063/PT.3.3167

Cronin, B. (2014). Scholars and Scripts, Spoors and Scores. In B. Cronin & C. R. Sugimoto (Eds.), *Beyond Bibliometrics: Harnessing Multidimensional Indicators of Scholarly Impact*. https://doi.org/10.7551/mitpress/9445.001.0001

Cronin, B., & Sugimoto, C. R. (Eds.). (2014). Beyond Bibliometrics.

Dehnel, M. P. (2011). Industrial Beamline Technologies And Approaches. *AIP Conference Proceedings*, *1336*, 21–25. https://doi.org/10.1063/1.3586049

Desai, J., Watson, D., Wang, V., Taddeo, M., & Floridi, L. (2022). The epistemological foundations of data science: A critical review. *Synthese*, *200*(6), 469. https://doi.org/10.1007/s11229-022-03933-2

Diamond Light Source. (2022). *Inside the Machine*. https://www.diamond.ac.uk/Home/About/FAQs/Inside-the-Machine.html

D'Ippolito, B., & Rüling, C.-C. (2019). Research collaboration in Large Scale Research Infrastructures: Collaboration types and policy implications. *Research Policy*, *48*(5), 1282–1296. https://doi.org/10.1016/j.respol.2019.01.011

D'Ippolito, B., & Rüling, C.-C. (2020). Keeping a Research Infrastructure alive: Material, social, and political work at the Institut Laue-Langevin. In *Big Science and Research Infrastructures in Europe* (pp. 232–250). Edward Elgar Publishing. https://doi.org/10.4337/9781839100017.00017

Donoho, D. (2017). 50 Years of Data Science. *Journal of Computational and Graphical Statistics*, *26*(4), 745–766. https://doi.org/10.1080/10618600.2017.1384734

EPN-Campus. (2014). *EPN-Campus: Joint ILL-ESRF Library*. *Available at* http://www.epn-campus.eu/library/joint-ill-esrf-library/

*ESRF Highlights 2002*. (2003). *Available at* https://www.esrf.fr/UsersAndScience/Publications/Highlights/2002

*ESRF Highlights 2004*. (2005). *Available at* https://www.esrf.fr/UsersAndScience/Publications/Highlights/2004

*ESRF Highlights 2022*. (2023). *Available at* https://www.esrf.fr/UsersAndScience/Publications/Highlights/esrf-highlights-2022.html

European Strategy Forum on Research Infrastructures. (2018). *Innovation-oriented cooperation of research infrastructures*. J. Moulin, M. Donzelli, M. Carpineti, & P. Dell'Arme (Eds.), Vol. 3. Dipartimento di Fisica - Università degli Studi di Milano.

European Synchrotron Radiation Facility. (n.d.-a). *A Brilliant Light for Science*. https://www.esrf.fr/about/information-material/brochures

European Synchrotron Radiation Facility. (n.d.-b). *About Us*. Retrieved 12 October 2023, from https://www.esrf.fr/about

European Synchrotron Radiation Facility. (n.d.-c). *Diamonds*. Retrieved 28 September 2023, from https://www.esrf.fr/UsersAndScience/Experiments/MX/About_our_beamlines/Beamline_Components/Diamonds

European Synchrotron Radiation Facility. (n.d.-d). *How does the ESRF work?* Retrieved 1 October 2023, from https://www.esrf.fr/home/education/what-is-the-esrf/how-does-the-esrf-work.html

European Synchrotron Radiation Facility. (n.d.-e). *ID23 Gemini: Technical Overview*. Retrieved 28 September 2023, from https://www.esrf.fr/UsersAndScience/Experiments/MX/About_our_beamlines/ID23-1/TechnicalOverview

European Synchrotron Radiation Facility. (n.d.-f). *Medium Term Scientific Programme for the period 2002 to 2006* (MSTP2002-2006). *Available at* https://www.esrf.fr/files/live/sites/www/files/UsersAndScience/Publications/mtspfiles/mtsp2002-2006.pdf

European Synchrotron Radiation Facility. (n.d.-g). *Medium-term Scientific Program (2001-2005)* (ESRF/SAC/41-03). *Available at* https://www.esrf.fr/files/live/sites/www/files/UsersAndScience/Publications/MTSP2001-2005/mtsp2001-2005.pdf

European Synchrotron Radiation Facility. (n.d.-h). *Support and Infrastructure*. Retrieved 12 June 2023, from https://www.esrf.fr/home/UsersAndScience/support-and-infrastructure.html

European Synchrotron Radiation Facility. (n.d.i). *ID19—Microtomography beamline*. https://www.esrf.fr/home/UsersAndScience/Experiments/StructMaterials/ID19.html

Fabre, R., Egret, D., Schöpfel, J., & Azeroual, O. (2021). Evaluating the scientific impact of research infrastructures: The role of current research information systems. *Quantitative Science Studies*, *2*(1), 42–64. https://doi.org/10.1162/qss_a_00111

Fecher, B., Kahn, R., Sokolovska, N., Völker, T., & Nebe, P. (2021). Making a Research Infrastructure: Conditions and Strategies to Transform a Service into an Infrastructure. *Science and Public Policy*, *48*(4), 499–507. https://doi.org/10.1093/scipol/scab026

Florio, M., Forte, S., & Sirtori, E. (2016). Forecasting the socio-economic impact of the Large Hadron Collider: A cost–benefit analysis to 2025 and beyond. *Technological Forecasting and Social Change*, *112*, 38–53. https://doi.org/10.1016/j.techfore.2016.03.007

Flot, D., Mairs, T., Giraud, T., Guijarro, M., Lesourd, M., Rey, V., van Brussel, D., Morawe, C., Borel, C., Hignette, O., Chavanne, J., Nurizzo, D., McSweeney, S., & Mitchell, E. (2010). The ID23-2 structural biology microfocus beamline at the ESRF. *Journal of Synchrotron Radiation*, *17*(Pt 1), 107–118. https://doi.org/10.1107/S0909049509041168

Frenken, K., Hardeman, S., & Hoekman, J. (2009). Spatial scientometrics: Towards a cumulative research program. *Journal of Informetrics*, *3*(3), 222–232. https://doi.org/10.1016/j.joi.2009.03.005

Frenken, K., & Hoekman, J. (2014). Spatial Scientometrics and Scholarly Impact: A Review of Recent Studies, Tools, and Methods. In Y. Ding, R. Rousseau, & D. Wolfram (Eds.), *Measuring Scholarly Impact* (pp. 127–146). Springer International Publishing. https://doi.org/10.1007/978-3-319-10377-8_6

Gabadinho, J., Hall, D., Leonard, G., Gordon, E., Monaco, S., & Thibault, X. (2008). Remote Access Experiments at the Macromolecular Crystallography Beamlines of the ESRF. *Synchrotron Radiation News*, *21*(5), 24–29. https://doi.org/10.1080/08940880802406075

Galison, P. (1997). *Image and Logic: A Material Culture of Microphysics*. University of Chicago Press.

Galison, P. (2010). Trading with the Enemy. In M. E. Gorman (Ed.), *Trading Zones and International Expertise: Creating New Kinds of Collaboration* (pp. 25–52). The MIT Press.

Galison, P., & Hevly, B. (Eds.). (1992). *Big Science: The Growth of Large-Scale Research.* Stanford University Press.

Gao, S. (2020). Spatial Scientometrics. In L. A. Schintler & C. L. McNeely (Eds.), *Encyclopedia of Big Data* (pp. 1–3). Springer International Publishing. https://doi.org/10.1007/978-3-319-32001-4_493-1

Garfield, E. (1955). Citation Indexes for Science: A New Dimension in Documentation through Association of Ideas. *Science*, *122*(3159), 108–111. https://doi.org/10.1126/science.122.3159.108

Gläser, J., & Laudel, G. (2001). Integrating Scientometric Indicators into Sociological Studies: Methodical and Methodological Problems. *Scientometrics*, *52*(3), 411–434. https://doi.org/10.1023/A:1014243832084

Goffman, W., & Newill, V. A. (1964). Generalization of Epidemic Theory: An Application to the Transmission of Ideas. *Nature*, *204*(4955). https://doi.org/10.1038/204225a0

Granovetter, M. S. (1973). The Strength of Weak Ties. *American Journal of Sociology*, *78*(6),

1360–1380. https://doi.org/10.1086/225469
Hallonsten, O. (2009). *Small science on big machines: Politics and practices of synchrotron radiation laboratories*. Lund University.
Hallonsten, O. (Ed.). (2012). *In pursuit of a promise: Perspectives on the political process to establish the European Spallation Source (ESS) in Lund, Sweden* (p. 234). KTH, History of Science, Technology and Environment.
Hallonsten, O. (2013). Introducing 'facilitymetrics': A first review and analysis of commonly used measures of scientific leadership among synchrotron radiation facilities worldwide. *Scientometrics*, *96*(2), 497–513. https://doi.org/10.1007/s11192-012-0945-9
Hallonsten, O. (2014). How expensive is Big Science? Consequences of using simple publication counts in performance assessment of large scientific facilities. *Scientometrics*, *100*(2), 483–496. https://doi.org/10.1007/s11192-014-1249-z
Hallonsten, O. (2015). The Parasites: Synchrotron Radiation at SLAC, 1972–1992. *Historical Studies in the Natural Sciences*, *45*(2), 217–272. https://doi.org/10.1525/hsns.2015.45.2.217
Hallonsten, O. (2016a). *Big Science Transformed: Science, Politics and Organization in Europe and the United States*. Palgrave Macmillan. https://doi.org/10.1007/978-3-319-32738-9
Hallonsten, O. (2016b). Use and productivity of contemporary, multidisciplinary Big Science. *Research Evaluation*, *25*(4), 486–495. https://doi.org/10.1093/reseval/rvw019
Hallonsten, O. (2020a). Research Infrastructures in Europe: The Hype and the Field. *European Review*, *28*(4), 617–635. https://doi.org/10.1017/S1062798720000095
Hallonsten, O. (2020b). *The Campaign: How a European Big Science facility ended up on the peripheral farmlands of southern Sweden*. Arkiv förlag.
Hallonsten, O., & Cramer, K. C. (2020). Big Science and Research Infrastructures in Europe: Conclusions and outlook. In *Big Science and Research Infrastructures in Europe* (pp. 251–257). Edward Elgar Publishing. https://doi.org/10.4337/9781839100017.00018
Hallonsten, O., Eriksson, H., & Collsiöö, A. (2020). The role of research infrastructures in innovation systems: The case of Swedish participation in the Halden Reactor Project. In O. Hallonsten & K. C. Cramer (Eds.), *Big Science and Research Infrastructures in Europe* (pp. 177–197). Edward Elgar Publishing. https://doi.org/10.4337/9781839100017.00014
Hallonsten, O., & Heinze, T. (2012). Institutional persistence through gradual organizational adaptation: Analysis of national laboratories in the USA and Germany. *Science and Public Policy*, *39*(4), 450–463. https://doi.org/10.1093/scipol/scs047
Hallonsten, O., & Heinze, T. (2015). Formation and expansion of a new organizational field in experimental science. *Science and Public Policy*, *42*(6), 841–854. https://doi.org/10.1093/scipol/scv013
Hammarfelt, B., & Åström, F. (2020). An Ecology of Measures and Indicators: Bibliometrics in Resource Allocation. In R. Ball (Ed.), *Handbook Bibliometrics* (pp. 291–298). De Gruyter. https://doi.org/10.1515/9783110646610
Heidler, R., & Hallonsten, O. (2015). Qualifying the performance evaluation of Big Science beyond productivity, impact and costs. *Scientometrics*, *104*(1), 295–312. https://doi.

org/10.1007/s11192-015-1577-7

Heinis, D., Carballedo, A., Colldelram, C., Cuní, G., Vidal, N. V., Matilla, Ó., Marcos, J., Sánchez, A., Casas, J., Nicolàs, J., Barrière, N., Collon, M. J., Vacanti, G., Handick, E., Müller, P., Krumrey, M., Ferreira, I., & Bavdaz, M. (2021). X-ray facility for the characterization of the Athena mirror modules at the ALBA synchrotron. *International Conference on Space Optics — ICSO 2020*, *11852*, 867–877. https://doi.org/10.1117/12.2599350

Heinze, T., & Hallonsten, O. (2017). The reinvention of the SLAC National Accelerator Laboratory, 1992–2012. *History and Technology*, *33*(3), 300–332. https://doi.org/10.1080/07341512.2018.1449711

Heinze, T., Hallonsten, O., & Heinecke, S. (2015). From Periphery to Center: Synchrotron Radiation at DESY : Part I: 1962–1977. *Historical Studies in the Natural Sciences*, *45*(3), 447–492.

Heinze, T., Hallonsten, O., & Heinecke, S. (2017). Turning the Ship: The Transformation of DESY, 1993–2009. *Physics in Perspective*, *19*(4), 424–451. https://doi.org/10.1007/s00016-017-0209-4

Heinze, T., & Jappe, A. (2020). Quantitative science studies should be framed with middle-range theories and concepts from the social sciences. *Quantitative Science Studies*, *1*(3), 983–992. https://doi.org/10.1162/qss_a_00059

Herfindahl, O. C. (1950). *Concentration in the steel industry*. Columbia University, New York.

Hicks, D., Wouters, P., Waltman, L., de Rijcke, S., & Rafols, I. (2015). Bibliometrics: The Leiden Manifesto for research metrics. *Nature*, *520*(7548). https://doi.org/10.1038/520429a

Hiltzik, M. (2015). *Big Science: Ernest Lawrence and the Invention that Launched the Military-Industrial Complex*. Simon and Schuster.

Hoddeson, L., Kolb, A. W., & Westfall, C. (2011). *Fermilab: Physics, the Frontier, and Megascience*. University of Chicago Press.

Hood, W. W., & Wilson, C. S. (2001). The Literature of Bibliometrics, Scientometrics, and Informetrics. *Scientometrics*, *52*(2), 291. https://doi.org/10.1023/A:1017919924342

Joerges, B., & Shinn, T. (Eds.). (2001). *Instrumentation Between Science, State and Industry*. Springer Netherlands. https://doi.org/10.1007/978-94-010-9032-2

Kessler, M. M. (1963). Bibliographic coupling between scientific papers. *American Documentation*, *14*(1), 10–25. https://doi.org/10.1002/asi.5090140103

Latour, B., & Woolgar, S. (1986). *Laboratory Life*.

Lauto, G., & Valentin, F. (2013). How Large-Scale Research Facilities Connect to Global Research. *Review of Policy Research*, *30*(4), 381–408. https://doi.org/10.1111/ropr.12027

Leigh Star, S. (2010). This is Not a Boundary Object: Reflections on the Origin of a Concept. *Science, Technology, & Human Values*, *35*(5), 601–617. https://doi.org/10.1177/0162243910377624

Leonardi, P. M. (2012). Materiality, Sociomateriality, and Socio-Technical Systems: What Do These Terms Mean? How are They Related? Do We Need Them? In P. M. Leonardi, B. A. Nardi, & J. Kallinikos (Eds.), *Materiality and Organizing: Social Interaction in a Technological World* (pp. 25–48). Oxford University Press.

Leydesdorff, L., Ràfols, I., & Milojević, S. (2020). Bridging the divide between qualitative and quantitative science studies. *Quantitative Science Studies*, *1*(3), 918–926. https://doi.org/10.1162/qss_e_00061

Lotka, A. J. (1926). The frequency distribution of scientific productivity. *Journal of the Washington Academy of Sciences*, *16*(12), 317–323.

Lucio-Arias, D., & Scharnhorst, A. (2012). Mathematical Approaches to Modeling Science from an Algorithmic-Historiography Perspective. In A. Scharnhorst, K. Börner, & P. van den Besselaar (Eds.), *Models of Science Dynamics: Encounters Between Complexity Theory and Information Sciences* (pp. 23–66). Springer. https://doi.org/10.1007/978-3-642-23068-4_2

Malbet-Monaco, S., Leonard, G. A., Mitchell, E. P., & Gordon, E. J. (2013). How the ESRF helps industry and how they help the ESRF. *Acta Crystallographica Section D: Biological Crystallography*, *69*(7). https://doi.org/10.1107/S0907444913001108

Mingers, J., & Leydesdorff, L. (2015). A review of theory and practice in scientometrics. *European Journal of Operational Research*, *246*(1), 1–19. https://doi.org/10.1016/j.ejor.2015.04.002

Moral-Munoz, J. A., López-Herrera, A. G., Herrera-Viedma, E., & Cobo, M. J. (2019). Science Mapping Analysis Software Tools: A Review. In W. Glänzel, H. F. Moed, U. Schmoch, & M. Thelwall (Eds.), *Springer Handbook of Science and Technology Indicators* (pp. 159–185). Springer International Publishing. https://doi.org/10.1007/978-3-030-02511-3_7

Nalimov, V. V., & Mulchenko, Z. M. (1971). *Measurement of Science. Study of the Development of Science as an Information Process,*. Air Force Systems Command, Wright-Patterson AFB, OH. Foreign Technology Div.

Paul-Hus, A., & Desrochers, N. (2019). Acknowledgements are not just thank you notes: A qualitative analysis of acknowledgements content in scientific articles and reviews published in 2015. *PLoS ONE*, *14*(12), e0226727. https://doi.org/10.1371/journal.pone.0226727

Petras, V. (2023). The identity of information science. *Journal of Documentation*, *ahead-of-print*(ahead-of-print). https://doi.org/10.1108/JD-04-2023-0074

Polanyi, M. (1962). *Personal knowledge: Towards a post-critical philosophy* (Repr. (with corr.)). Routledge & Kegan Paul.

Pomerantz, J. (2015a). Introduction. In *Metadata* (pp. 1–18). The MIT Press

Pomerantz, J. (2015b). Metadata. The MIT Press.

Price, D. J. D. (1963). *Little Science, Big Science*. Columbia University Press. https://doi.org/10.7312/pric91844

Price, D. J. D. (1965). Networks of Scientific Papers. *Science*, *149*(3683), 510–515. https://doi.org/10.1126/science.149.3683.510

Price, D. J. D. (1984). The science/technology relationship, the craft of experimental science, and policy for the improvement of high technology innovation. *Research Policy*, *13*(1), 3–20. https://doi.org/10.1016/0048-7333(84)90003-9

Raimondi, P. (2016). ESRF-EBS: The Extremely Brilliant Source Project. *Synchrotron Ra-*

*diation News*, *29*(6), 8–15. https://doi.org/10.1080/08940886.2016.1244462

Rekers, J. V., & Sandell, K. (Eds.). (2016). *New big science in focus: Perspectives on ESS and MAX IV* (Vol. 8). Lund Studies in Arts and Cultural Sciences.

Rhoades, S. A. (1993). The Herfindahl-Hirschman index. *Federal Reserve Bulletin*, *Mar*, 188–189.

Rift, G. (2023). *Beamlines*. MAX IV. https://www.maxiv.lu.se/beamlines-accelerators/beamlines/

Rosenberg, N. (1992). Scientific instrumentation and university research. *Research Policy*, *21*(4), 381–390. https://doi.org/10.1016/0048-7333(92)90035-3

Sawyer, S., Erickson, I., & Jarrahi, M. H. (2019). Infrastructural Competence. In J. Vertesi, D. Ribes, C. DiSalvo, Y. Loukissas, L. Forlano, D. K. Rosner, S. J. Jackson, & H. R. Shell (Eds.), *DigitalSTS* (pp. 267–279). Princeton University Press; JSTOR. https://doi.org/10.2307/j.ctvc77mp9.22

SESAME. (2020). *SESAME's Beamlines*. https://www.sesame.org.jo/beamlines

Shrum, W., Genuth, J., & Chompalov, I. (2007). *Structures of Scientific Collaboration*. The MIT Press.

Silva, F. S. V., Schulz, P. A., & Noyons, E. C. M. (2019). Co-authorship networks and research impact in large research facilities: Benchmarking internal reports and bibliometric databases. *Scientometrics*, *118*(1), 93–108. https://doi.org/10.1007/s11192-018-2967-4

Simpson, E. H. (1949). Measurement of Diversity. *Nature*, *163*(4148), 688–688. https://doi.org/10.1038/163688a0

Small, H. (1973). Co-citation in the scientific literature: A new measure of the relationship between two documents. *Journal of the American Society for Information Science*, *24*(4), 265–269. https://doi.org/10.1002/asi.4630240406

Smirnova, N., & Mayr, P. (2023). A comprehensive analysis of acknowledgement texts in Web of Science: A case study on four scientific domains. *Scientometrics*, *128*(1), 709–734. https://doi.org/10.1007/s11192-022-04554-9

Söderström, K. R. (2023a). Global reach, regional strength: Spatial patterns of a big science facility. Journal of the Association for Information Science and Technology, 74(9), 1140–1156. https://doi.org/10.1002/asi.24811

Söderström, K. R. (2023b). The structure and dynamics of instrument collaboration networks. Scientometrics, 128(6), 3581–3600. https://doi.org/10.1007/s11192-023-04658-w

Söderström, K. R., Åström, F., & Hallonsten, O. (2022). Generic instruments in a synchrotron radiation facility. Quantitative Science Studies, 3(2), 420–442. https://doi.org/10.1162/qss_a_00190

SOLEIL Synchrotron. (2016). *Beamlines*. https://www.synchrotron-soleil.fr/en/beamlines

Star, S. L., & Bowker, G. C. (2010). How to Infrastructure. In *Handbook of New Media: Social Shaping and Social Consequences of ICTs, Updated Student Edition* (pp. 230–245). SAGE Publications Ltd. https://doi.org/10.4135/9781446211304

Star, S. L., & Griesemer, J. R. (1989). Institutional Ecology, 'Translations' and Boundary Objects: Amateurs and Professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39. *Social Studies of Science*, *19*(3), 387–420.

Star, S. L., & Ruhleder, K. (1996). Steps Toward an Ecology of Infrastructure: Design and Access for Large Information Spaces. *Information Systems Research*, *7*(1), 111–134. https://doi.org/10.1287/isre.7.1.111

Sundin, O., Andersson, C., & Söderström, K. (2023). Controlling the machinery of knowledge: Google and access to COVID-19 grey-zone medicines. In S. Lundin, R. Liu, A. Smith, & E. Muller (Eds.), *Medicines Across Borders, Exploration of Grey Zones (In press)*. African Sun Media.

Sundin, O., & Haider, J. (2022). *Paradoxes of Media and Information Literacy: The Crisis of Information*. Routledge. https://doi.org/10.4324/9781003163237

Tague-Sutcliffe, J. (1992). An introduction to informetrics. *Information Processing & Management*, *28*(1), 1–3. https://doi.org/10.1016/0306-4573(92)90087-G

Trist, E. L., & Bamforth, K. W. (1951). Some Social and Psychological Consequences of the Longwall Method of Coal-Getting: An Examination of the Psychological Situation and Defences of a Work Group in Relation to the Social Structure and Technological Content of the Work System. *Human Relations*, *4*(1), 3–38. https://doi.org/10.1177/001872675100400101

Vemuri, V. (1978). Language of System Theory. In V. Vemuri (Ed.), *Modeling of Complex Systems* (pp. 25–66). Academic Press. https://doi.org/10.1016/B978-0-12-716550-9.50009-1

Venturini, T., Jensen, P., & Latour, B. (2015). Fill in the Gap. A New Alliance for Social and Natural Sciences. *Journal of Artificial Societies and Social Simulation*, *18*(2), 11.

Vertesi, J. (2019). Introduction. In J. Vertesi, D. Ribes, C. DiSalvo, Y. Loukissas, L. Forlano, D. K. Rosner, S. J. Jackson, & H. R. Shell (Eds.), *DigitalSTS* (pp. 263–266). Princeton University Press; JSTOR. https://doi.org/10.2307/j.ctvc77mp9.21

Virgin Galactic. (n.d.). *Galactic 01: Research Payloads*. Retrieved 20 July 2023, from https://www.virgingalactic.com/galactic-01-research-payloads-fact-sheet

Waltman, L., Larivière, V., Milojević, S., & Sugimoto, C. R. (2020). Opening science: The rebirth of a scholarly journal. *Quantitative Science Studies*, *1*(1), 1–3. https://doi.org/10.1162/qss_e_00025

Webb, C. (2020). *SpaceX-20 Experiments and Payloads | NASA*. https://www.nasa.gov/feature/space-life-and-physical-sciences-research-and-applications-spacex-20-experiments-and-payloads

Weinberg, A. M. (1961). Impact of Large-Scale Science on the United States: Big science is here to stay, but we have yet to make the hard financial and educational choices it imposes. *Science*, *134*(3473), 161–164. https://doi.org/10.1126/science.134.3473.161

Weinberg, A. M. (1968). *Reflections on Big Science*. MIT Press.

Weitkamp, T., Tafforeau, P., Boller, E., Cloetens, P., Valade, J., Bernard, P., Peyrin, F., Ludwig, W., Helfen, L., & Baruchel, J. (2010). Status and evolution of the ESRF beamline ID19. *AIP Conference Proceedings*, *1221*(1), 33–38. https://doi.org/10.1063/1.3399253

Westfall, C. (2003). Rethinking Big Science: Modest, Mezzo, Grand Science and the Development of the Bevalac, 1971–1993. *Isis*, *94*(1), 30–56. https://doi.org/10.1086/376098

Westfall, C. (2008). Retooling for the Future: Launching the Advanced Light Source at

Lawrence's Laboratory, 1980–1986. *Historical Studies in the Natural Sciences*, *38*(4), 569–609. https://doi.org/10.1525/hsns.2008.38.4.569

Westfall, C., & Hoddeson, L. (1996). Thinking Small in Big Science: The Founding of Fermilab, 1960-1972. *Technology and Culture*, *37*(3), 457–492. https://doi.org/10.2307/3107161

White, H., & McCain, K. (1989). Bibliometrics. *Annual Review of Information Science and Technology*, *24*, 119–186.

Wouters, P. (2014). The Citation: From Culture to Infrastructure. In B. Cronin & C. R. Sugimoto (Eds.), *Beyond Bibliometrics: Harnessing Multidimensional Indicators of Scholarly Impact*. https://doi.org/10.7551/mitpress/9445.001.0001

Wyatt, S., Milojević, S., Park, H., & Leydesdorff, L. (2015). Quantitative and Qualitative STS: The Intellectual and Practical Contributions of Scientometrics. *Available at SSRN: Https://Ssrn.Com/Abstract=2588336 or Http://Dx.Doi.Org/10.2139/Ssrn.2588336*.

Zemplén, G. Á. (2019). The contribution of laboratory studies, science studies and Science and Technology Studies (STS) to the understanding of scientific communication. In A. Leßmöllmann, M. Dascal, & T. Gloning (Eds.), *Science Communication* (Vol. 17, pp. 123–142). De Gruyter Mouton. https://doi.org/10.1515/9783110255522-006

# Lund Studies in Arts and Cultural Sciences

## Previously Published in the Series:

31. Kristofer Rolf Söderström 2023 *The dynamics of beamline configurations and user communities Quantitative studies of Big Science publications*
30. Carin Graminius 2023. *Research communication in the climate crisis Open letters and the mobilization of information.*
29. Andreas Helles Pedersen 2023. *Listening from within a Digital Music Archive. Metadata, sensibilities, and music histories in the Danish broadcasting corporation's music archive.*
28. Bruno Hamnell 2021. *Two Quests for Unity. John Dewey, R. G. Collingwood, and the Persistence of Idealism*.
27. Cecilia Andersson 2021. *Performing search. Search Engines and Mobile Devices in the Everyday Life of Young People*
26. Schmidt, Nora 2020. *The Privilege to Select Global Research System, European Academic Library Collections, and Decolonisation*.
25. Martin , Christopher 2020. *Shifting Gears. Automated Driving on the Eve of Autonomous Drive.*
24. Höög, Victoria, Kärrholm, Sara & Nilsson, Gabriella (red.) 2019. *Kultur X. 10-talet i kulturvetenskaplig belysning*.
23. Olsson Dahlquist, Lisa 2019. *Folkbildning för delaktighet. En studie om bibliotekets demokratiska uppdrag i en digital samtid*.
22. Jönsson, Jimmy 2019. *Den biologiska vändningen. Biologi och skogsvård, 1900–1940*.
21. Wiszmeg, Andréa 2019. *Cells in Culture, Cells in Suspense. Practices of Cultural Production in Foetal Cell Research.*
20. Hanell, Fredrik 2019. *Lärarstudenters digitala studievardag. Informationslitteracitet vid en förskollärarutbildning.*
19. Herd, Katarzyna 2018. *"We can make new history here". Rituals of producing history in Swedish football clubs*.
18. Qvarnström, Ludwig (ed.) 2018. *Swedish Art History. A Selection of Introductory Texts*.
17. Salomonsson, Karin (red.) 2018. *Mitt och ditt. Etnologiska perspektiv på ägandets kulturella betydelse*.

16. Bengtsen, Peter, Liljefors, Max & Petersén, Moa (red.) 2018. *Bild och natur. Tio konstvetenskapliga betraktelser*.
15. Cridland, Meghan 2017. *'May contain traces of'. An Ethnographic Study of Eating Communities and the Gluten Free Diet*.
14. Jönsson, Lars-Eric (red.) 2017. *Politiska projekt, osäkra kulturarv*.
13. Jönsson, Lars-Eric & Nilsson, Fredrik (red.) 2017. *Kulturhistoria. En etnologisk metodbok*.
12. Askander, Mikael 2017. *Poesier & kombinationer. Bruno K. Öijers intermediala poesi*.
11. Alftberg, Åsa, Apelmo, Elisabet & Hansson, Kristofer (red.) 2016. *Ljud tar plats. Funktionshinderperspektiv på ljudmiljöer*.
10. Arvidson, Mats 2016. *An Imaginary Musical Road Movie. Transmedial Semiotic Structures in Brad Mehldau's Concept Album Highway Rider*.
9. Brenthel, Adam 2016. *The Drowning World. The Visual Culture of Climate Change*.
8. Rekers, Josephine V. & Sandell, Kerstin (eds.) 2016. *New Big Science in Focus. Perspectives on ESS and MAX IV*.
7. Gunnarson, Martin 2016. *Please Be Patient. A Cultural Phenomenological Study of Haemodialysis and Kidney Transplantation Care*.
6. Lindh, Karolina 2015. *Breathing Life into a Standard. The configuration of resuscitation in practices of informing*.
5. Jönsson, Lars-Eric & Nilsson, Fredrik (red.) 2014. S*kratt som fastnar. Kulturella perspektiv på skratt och humor*.
4. Blaakilde, Anne Leonora & Nilsson, Gabriella (eds.) 2013. *Nordic Seniors on the Move. Mobility and Migration in Later Life*.
3. Carlsson, Hanna 2013. *Den nya stadens bibliotek. Om teknik, förnuft och känsla i gestaltningen av kunskaps- och upplevelsestadens folkbibliotek*.
2. Hagen, Niclas 2013. *Modern Genes. Body, Rationality and Ambivalence*.
1. Lysaght, Patricia (ed.) 2013. *The Return of Traditional Food. Proceedings of the 19th International Ethnological Food Research Conference*.

In the evolving landscape of Big Science's user-centric turn, this thesis provides insights into the role of beamlines at the European Synchrotron Radiation Facility (ESRF). Conceptualising beamlines as unique collaborative spaces, it employs quantitative techniques on scientific publication metadata to reveal how beamline configurations shape access, collaboration, and knowledge production between diverse user communities. Drawing from over 30,000 ESRF publications from 1994-2021, this thesis employs various sample sizes across four research articles to analyse the disciplinary, collaborative, geographical, and informal processes associated with beamlines. The results reveal that beamlines attract specialised and interdisciplinary user communities based on the technological configurations of their scientific instruments. Moreover, continuous upgrades to beamline instrumentation often led to increased collaboration and participation from diverse global user communities. Overall, the thesis emphasises beamlines as cross-disciplinary spaces fostering both specialised and interdisciplinary research. It highlights the significant, yet often overlooked, impact of beamline configurations on shaping user access and collaboration within the ESRF. Methodologically, it enriches the scientometrics toolbox through the novel application of computational techniques to publication metadata. Theoretically, it advances the understanding of user-oriented big science facilities as interconnected beamline infrastructures that adapt to serve evolving user communities.

Kristofer Rolf Söderström, Department of Arts and Cultural Sciences, Lund University. The dynamics of beamline configurations and user communities is his doctoral thesis in Information Studies.

LUND UNIVERSITY

LUND STUDIES IN ARTS AND CULTURAL SCIENCES
ISBN 978-91-987325-4-2
ISSN 2001-7529