



Ministry of Higher  
Education and Science

Danish Agency for Science  
and Higher Education

# The Danish Bibliometric Research Indicator (BFI)

**Lotte Faurbæk**

Head of Section

# Agenda

## 1. Background

## 2. Data types

- **Publication types**
- **External data sources**

## 3. Harvest procedure and protocol

- **Harvest flow**
- **Harvest protocol**

## 4. Automatic validation

- **7 Screening rules**
- **8 Classification rules**

## 5. Manual validation

## 6. Some numbers from BFI



# Background

BFI is a part of the performance based funding model for distribution of new basic funding to Danish universities

➤ **Before 2009:**

- ✓ **50-40-10-model used to distribute new funding:**
  - Education grants, weight 50 %
  - External funding, weight 40 %
  - Number of ph.d.'s, weight 10 %

➤ **After 2009:**

- ✓ **Political agreement June 2009: The Bibliometric Research Indicator becomes part of the performance based funding model**
- ✓ **Gradual implementation over 3 years (10%, 15% og 25%)**

➤ **Today:**

- Education grants, weight 45 %
- External funding, weight 20 %
- BFI, weight 25 %
- Number of ph.d.'s, weight 10 %



# Publication types i BFI

**Only peer reviewed research publications elicite points in BFI!**

- **Scientific monographs (publisher channel)**
- **Scientific monographs in book series**
- **Scientific articles i series (journals and conference series)**
- **Scientific chapters in a book in a book series**
- **Scientific book chapters (publisher channel)**
- **Doctoral dissertations**
- **Patents**



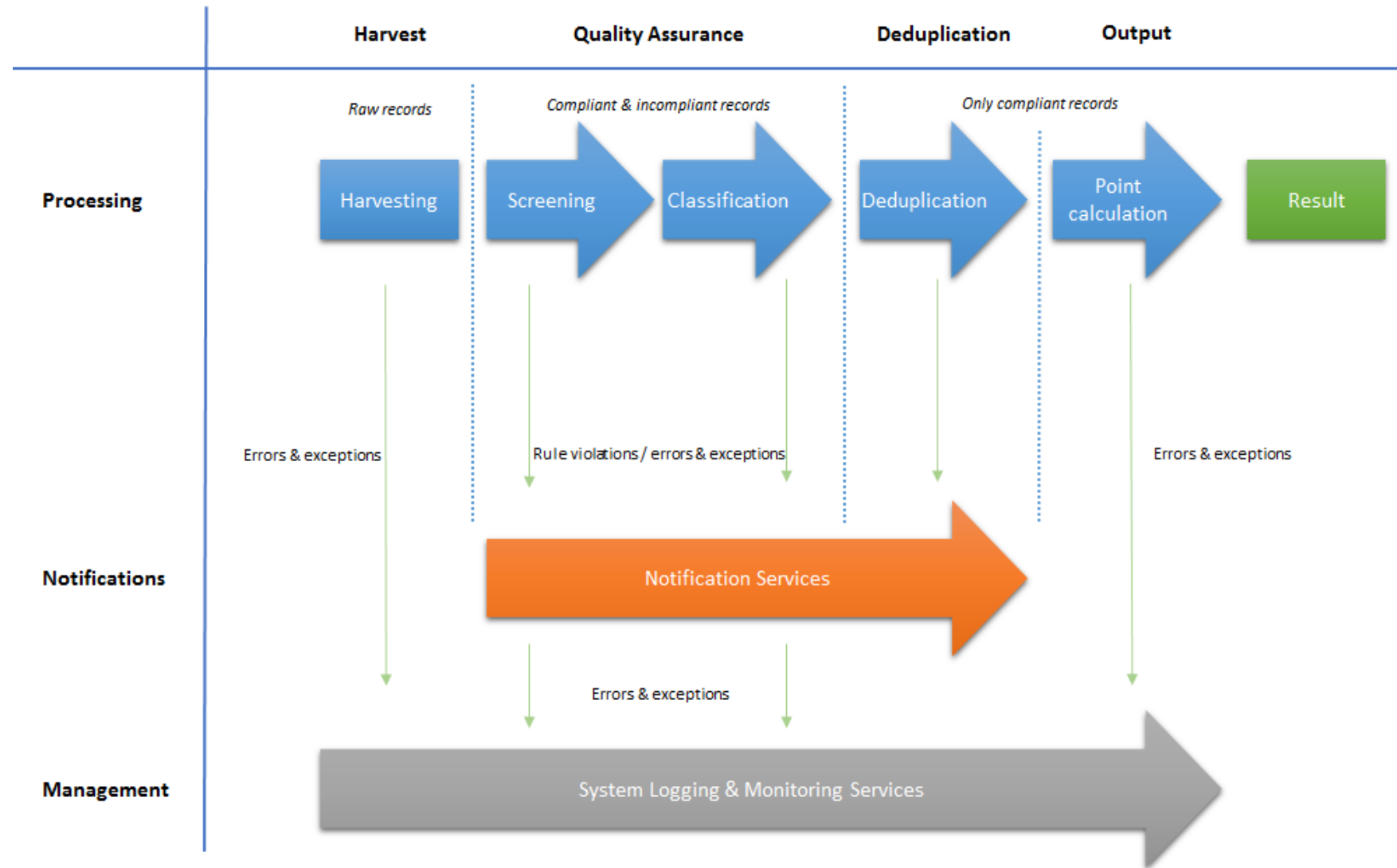
# External data sources in the BFI data base

## Import of external data

- **Ulrichs periodical directory**
- **Journal Citation Report (Clarivate)**
- **Sherpa/Romeo**
- **BFI-specific data from previous years (levels, expert panels etc.)**
- **Finnish and Norwegian levels**
- **Relative impact factors**
- **World Production**



# Harvest flow



# Harvest protocol

Only metadata are harvested from the Danish universities

- **Harvest protocol: OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting)**
- **DDF-MXD-format**
- **XML-records**
- **Validation against schema**
- **XML-records registered the previous year (BFI analysis year) and registered the current year are harvested (e.g. this year we have harvested 2017 and 2018 publications)**
- **Every university are harvested twice a night (changes since last harvest)**



# Automatic validation

## Screening rules

Harvested publications pass through 7 automatic checks in sequence!

1. XML format must be readable and compliant with DDF-MXD schema
2. The reporting year must be a valid number and be equal to analysis year
3. Publication year must be a valid number (must consist of 4 digits)
4. No more than +/- one year between annual reporting and publication year
5. The publication must contain at least one local author
6. The publication must be attached to a main research area (HUM, SOC, SCI/TECH and HEALTH)
7. The publication must not have gotten points in previous years (blacklisted)





# Automatic validation

## Classification rules I

Then, harvested publications pass through 8 classification rules in sequence!

1. The publication must be scientific (must have the MXD code “sci”)
2. Status of the publication must be published (must have the MXD code “p”)
3. The publication (except patents and doctoral dissertations) must be peer-reviewed (must have the MXD code “pr”)
4. The document type can be found in the point table (article, review article, conference article, book, report, book chapter, chapter in book in a book series, chapter in report, conference article in proceeding, doctoral dissertation, patent)



# Automatic validation

## Classification rules II

5. For articles, review articles , conference articles, chapters in book in a book series and chapters in a report in a book series, the publication must have a publication channel on the BFI list of series
6. For books, book chapters, reports and report chapters, the publication must have a publication channel on the BFI list of publishers
7. For doctoral dissertations and patents, the publication must elicit point regardless of publication channel
8. If the publication does not comply with at least one of the above mentioned classification rules, the publication will be classified as "none"

Here after the automatic deduplication starts



# Manual validation

## Deduplication

6.000–7.000 duplicates pr. year

### 1. Deduplication:

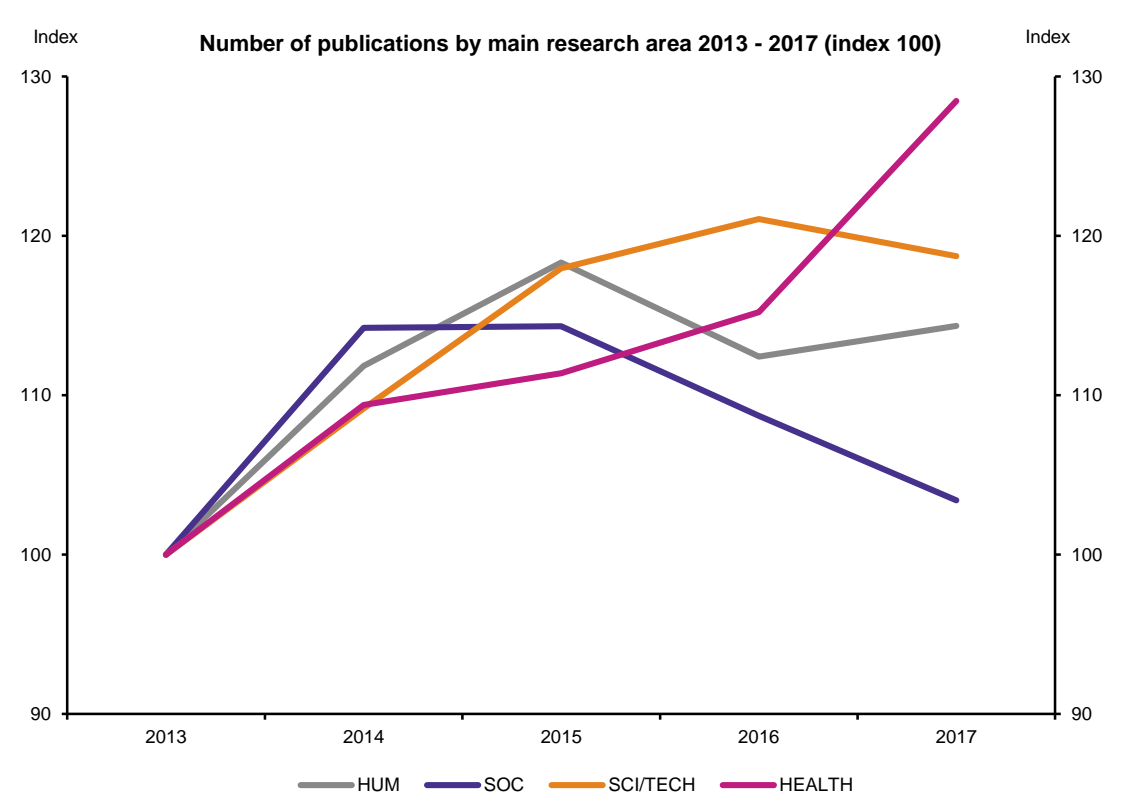
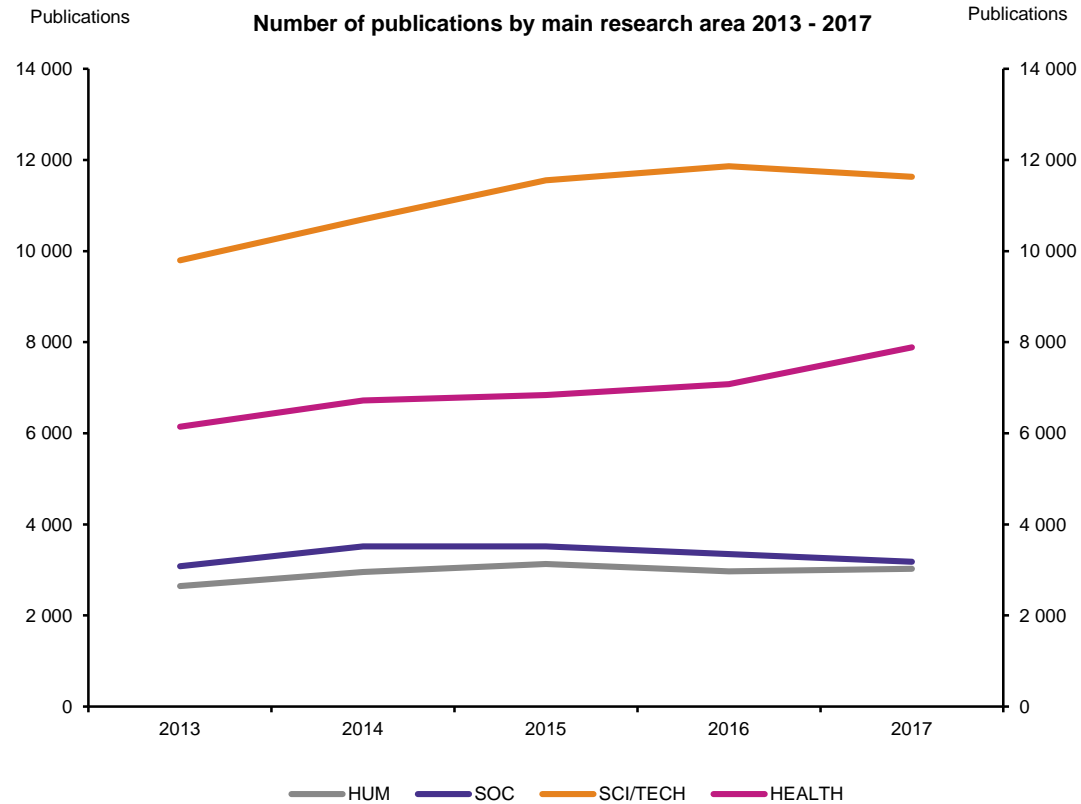
- **KVM (quality manager) takes care of duplicates and identifies conflicts**
- **KVM sends conflicts to the university's validators**

### 2. Conflicts:

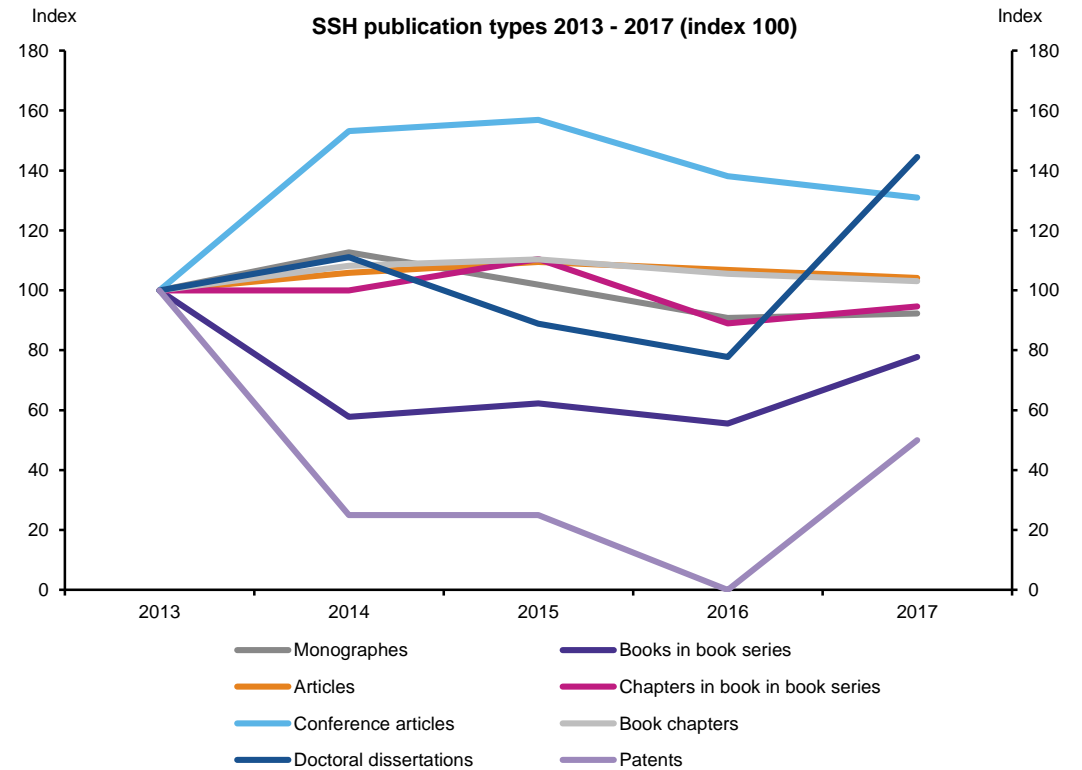
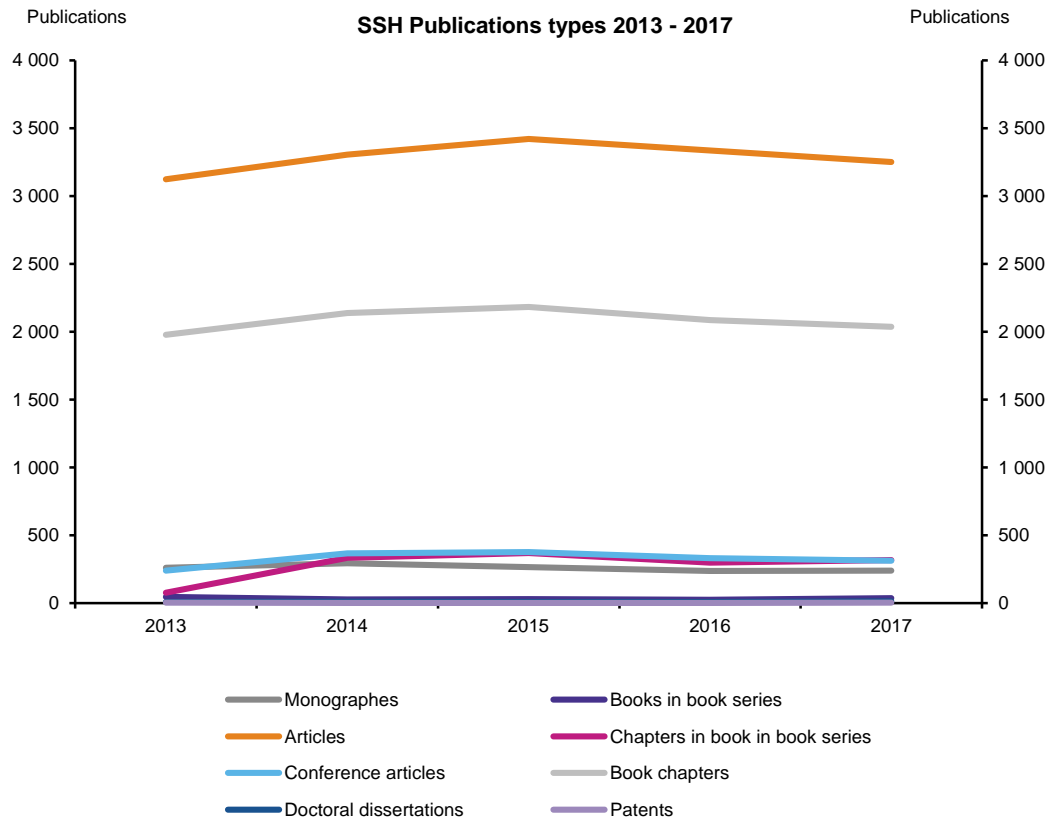
- **2 kinds of conflicts:**
  - ❖ **Author conflicts**
  - ❖ **Classification conflicts**
- **The system notifies validators at the relevant universities**
- **Validators edit and agree on the authoritative record in BFI**



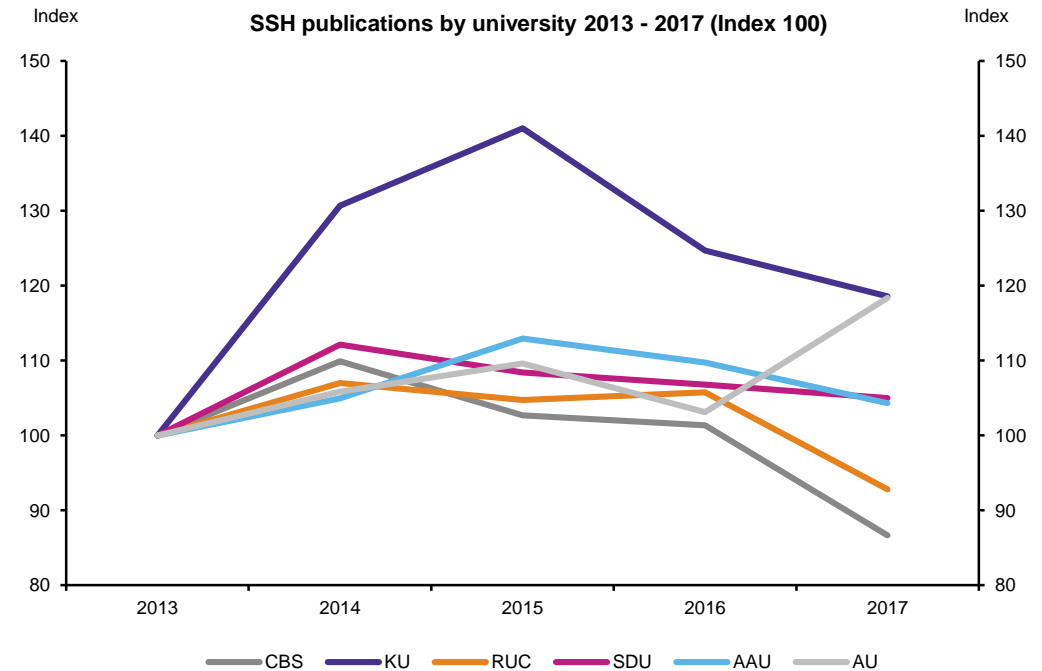
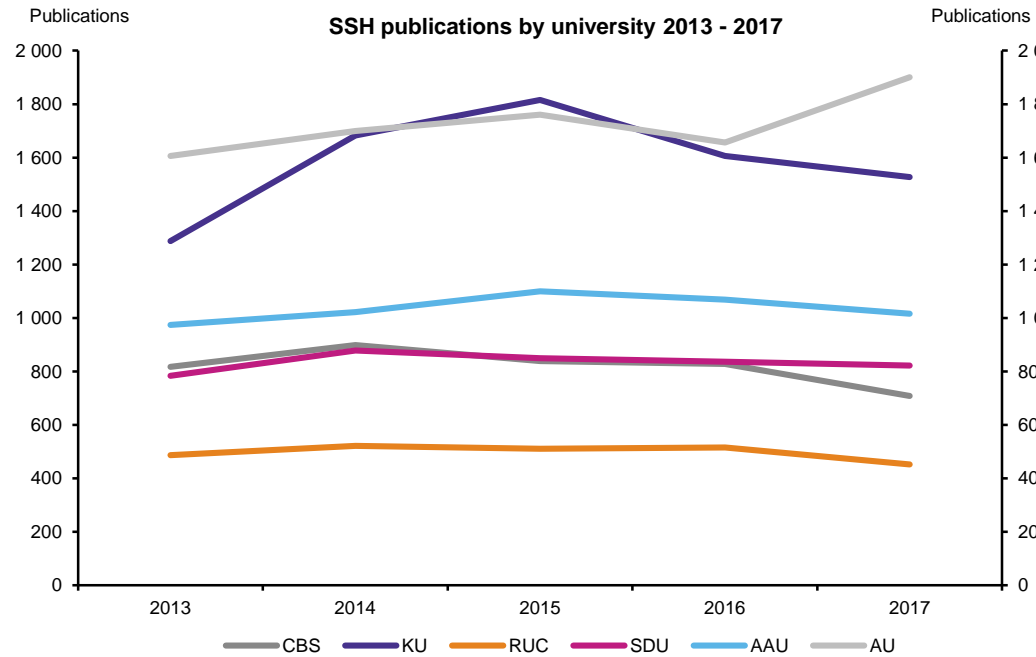
# Publications by main research area



# SSH publication types

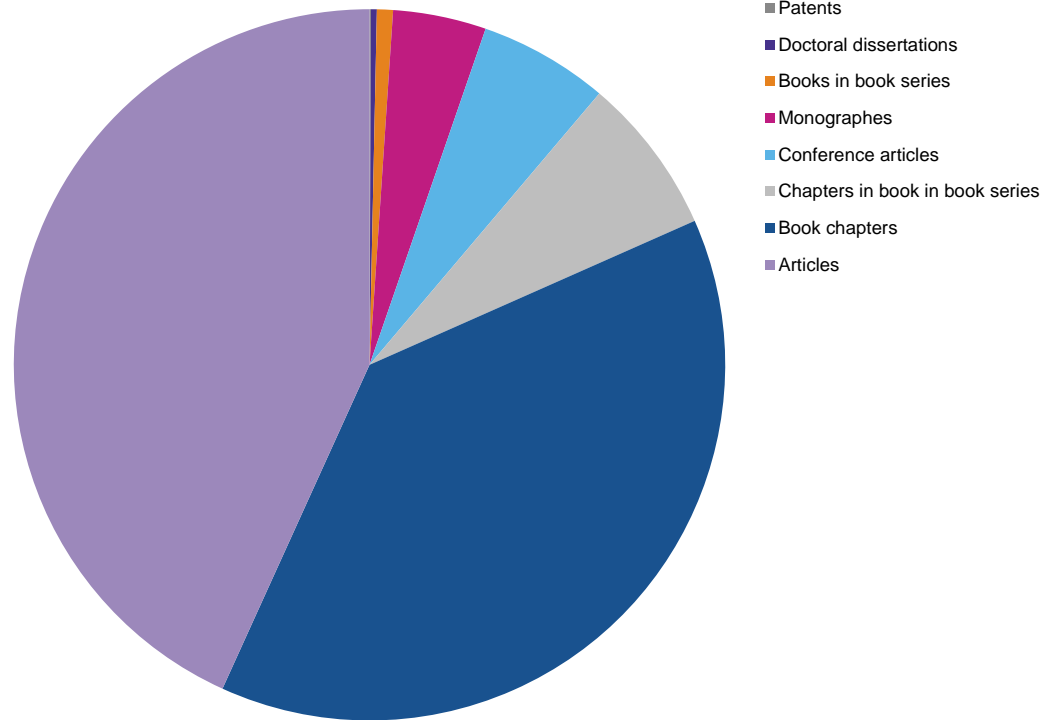


# SSH publication types by university

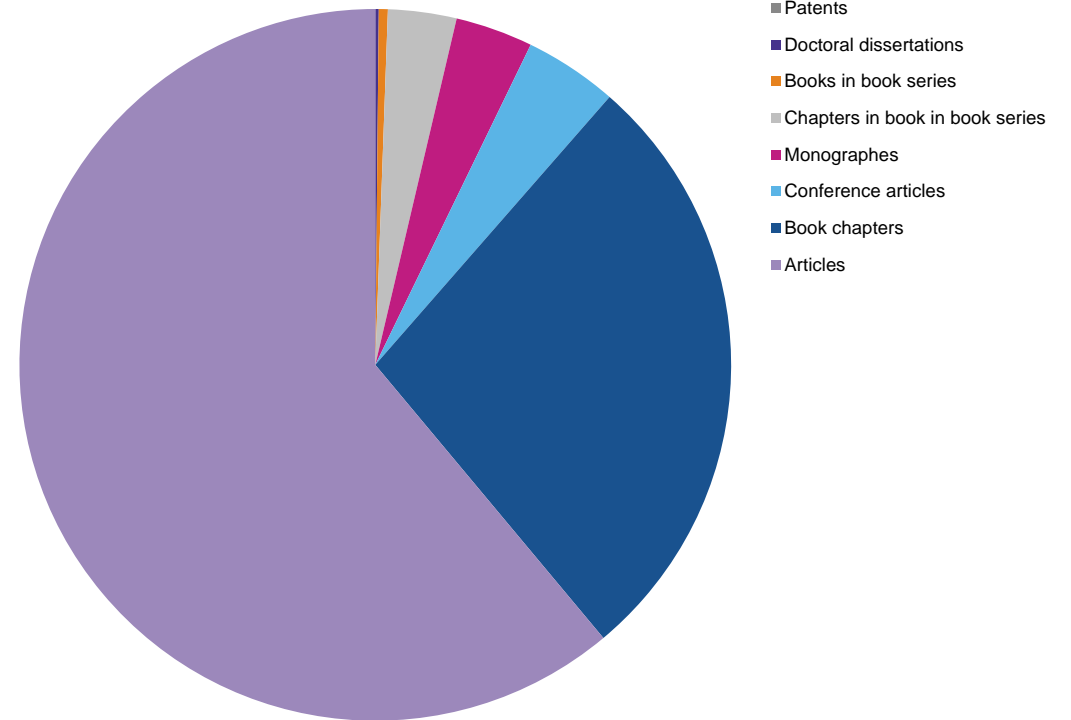


# HUM and SOC publication types 2017 (%)

HUM publication types 2017



SOC publication types 2017



**Thank you for the attention!**