Information scientists: between editors and data centers

Marianne Brouty

F. Woelfel, S. Borde, C. Bruneau, C. Brunet, G. Chassagnard, H. Claude, P. Dubois, A. Eisele, F. Genova, S. Lesteven, M. Neuville, F. Ochsenbein, E. Perret, P. Vannier, P. Vonflie

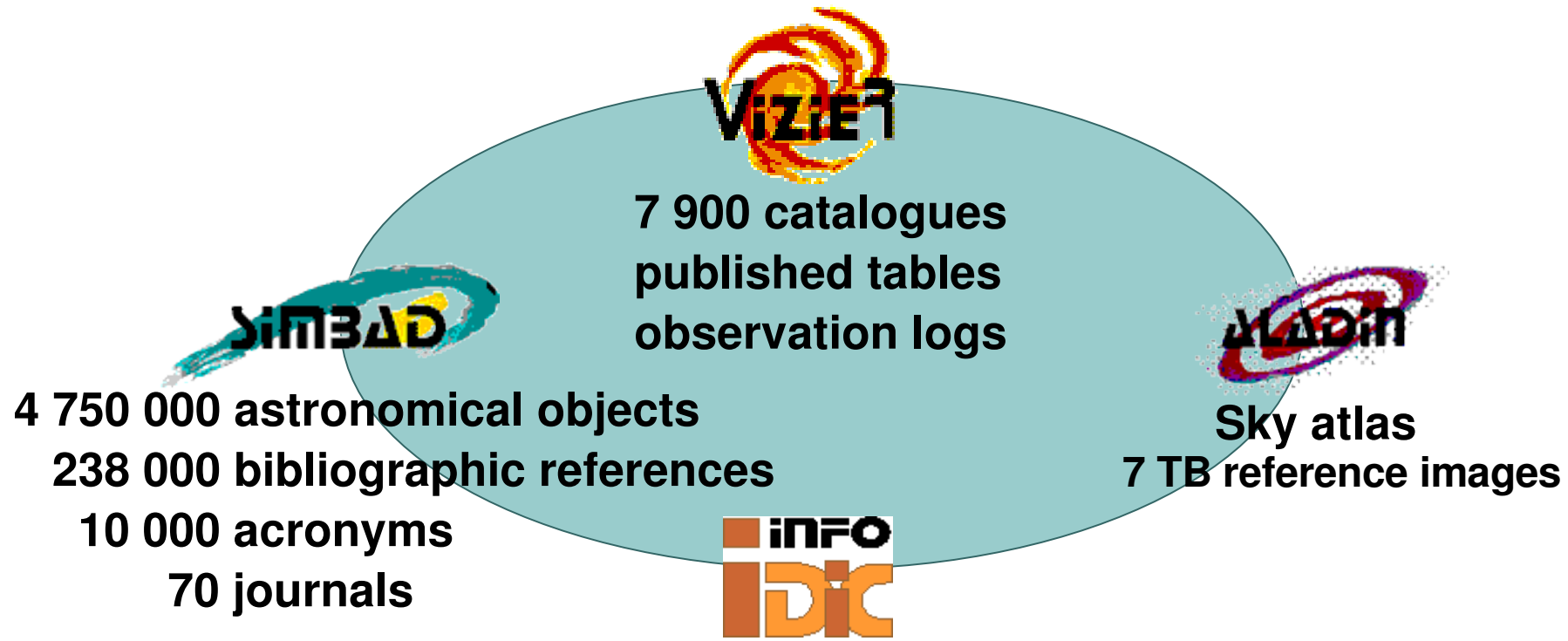
LISA VI - Pune, India
February 14-17, 2010

CDS
Centre de Données astronomiques de Strasbourg

Centre de Données de Strasbourg

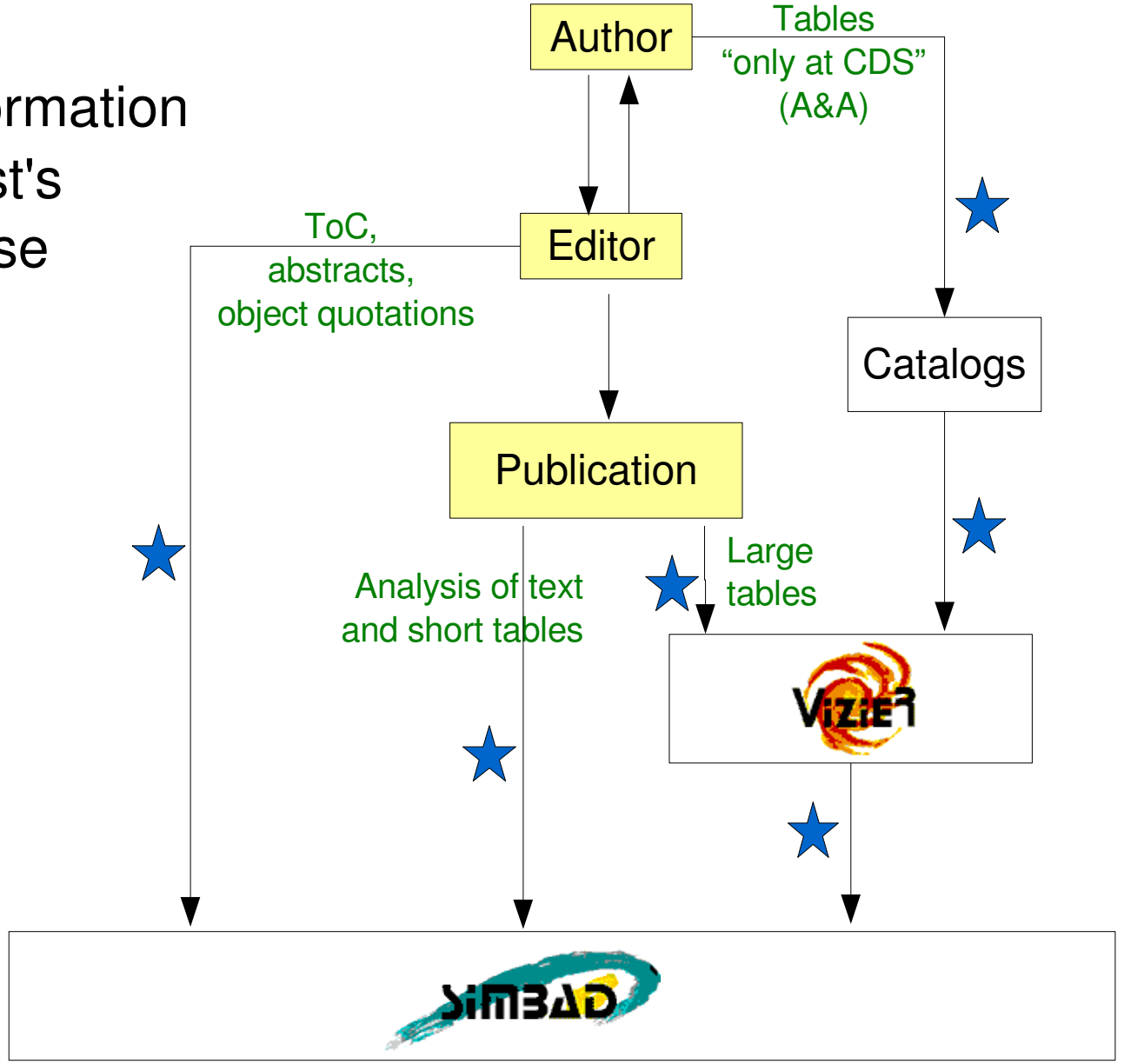
- Founded in 1972
- Ongoing collaboration with other data centers
- On the Internet since 1991
- Involved in the publication of electronic tables of *Astronomy & Astrophysics* since 1993
- Develops and maintains on-line services

The CDS on-line services



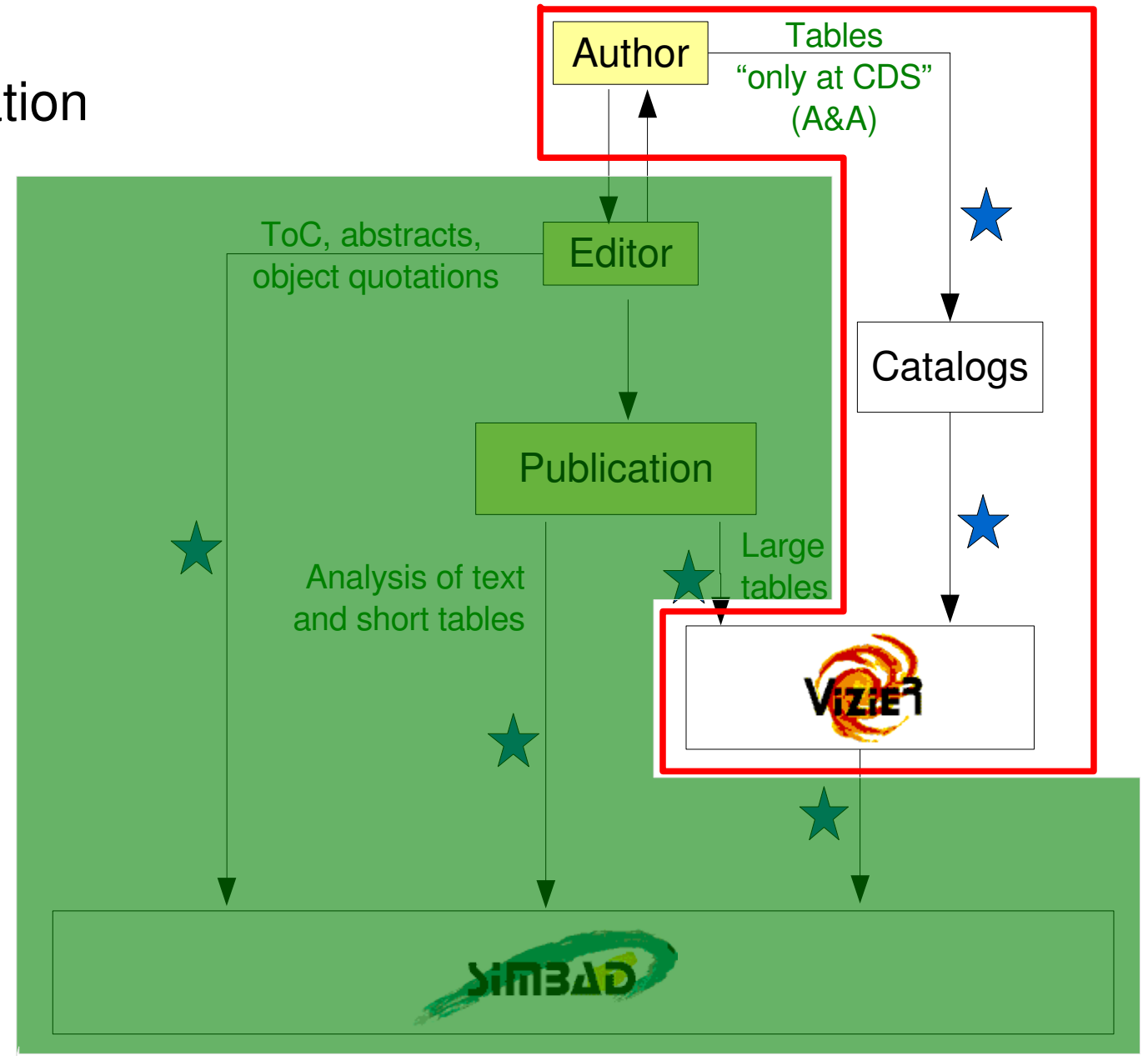
Life cycle of an article

- ★ : information scientist's expertise



Electronic tables (1)

- ★ : information scientist's expertise

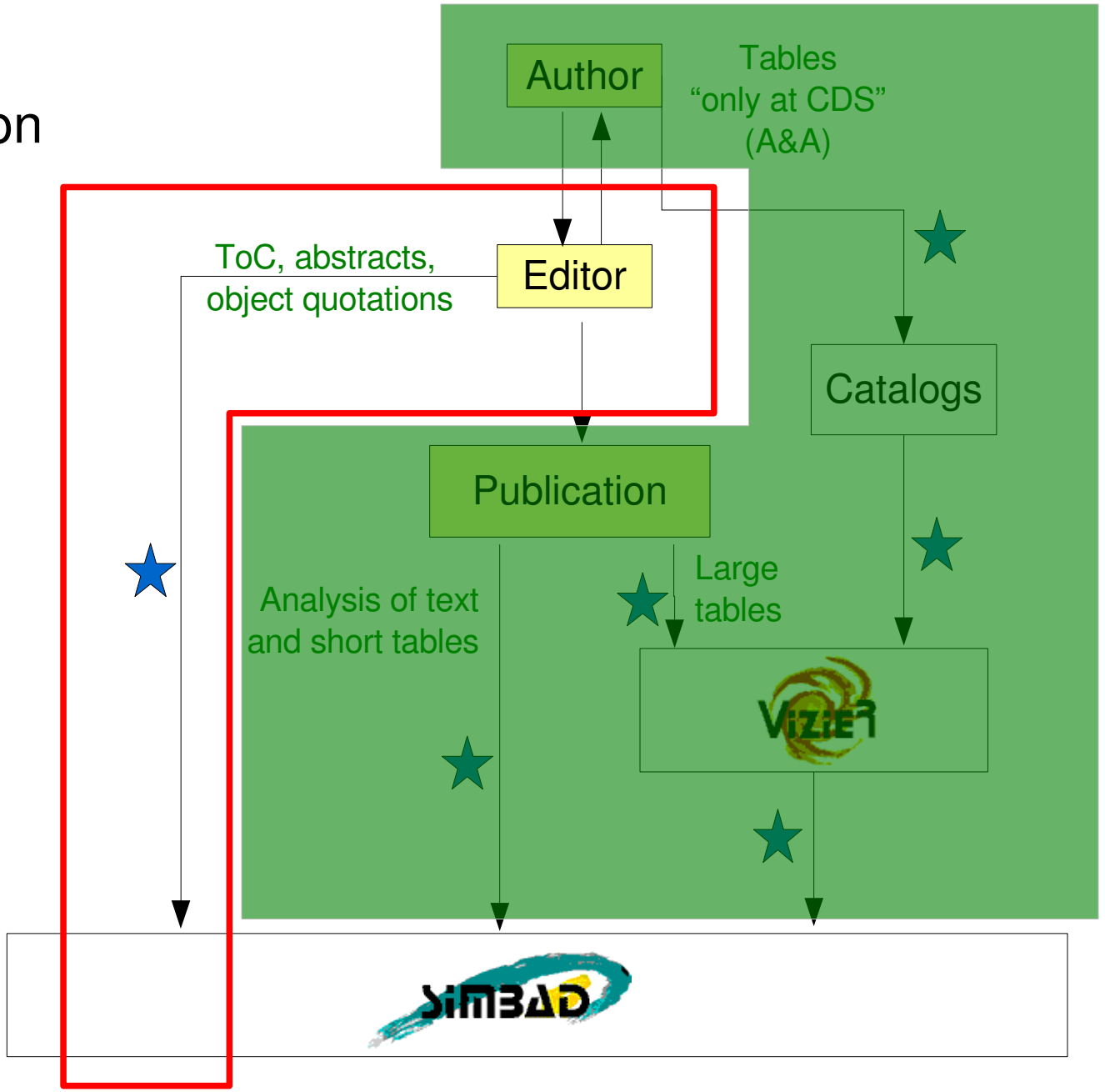


Electronic tables (2)

- A&A: part of the publication process
  - Scientific editor
    - tables "only at CDS"
    - at editor's request:
      tables+description sent directly to the CDS by authors
  - CDS
    - descriptions homogenized and standardized
    - check (data consistency, may imply interactions with the authors)
    - become public upon publication
    - accessible via VizieR

Tables of contents (1)

- ★ : information scientist's expertise

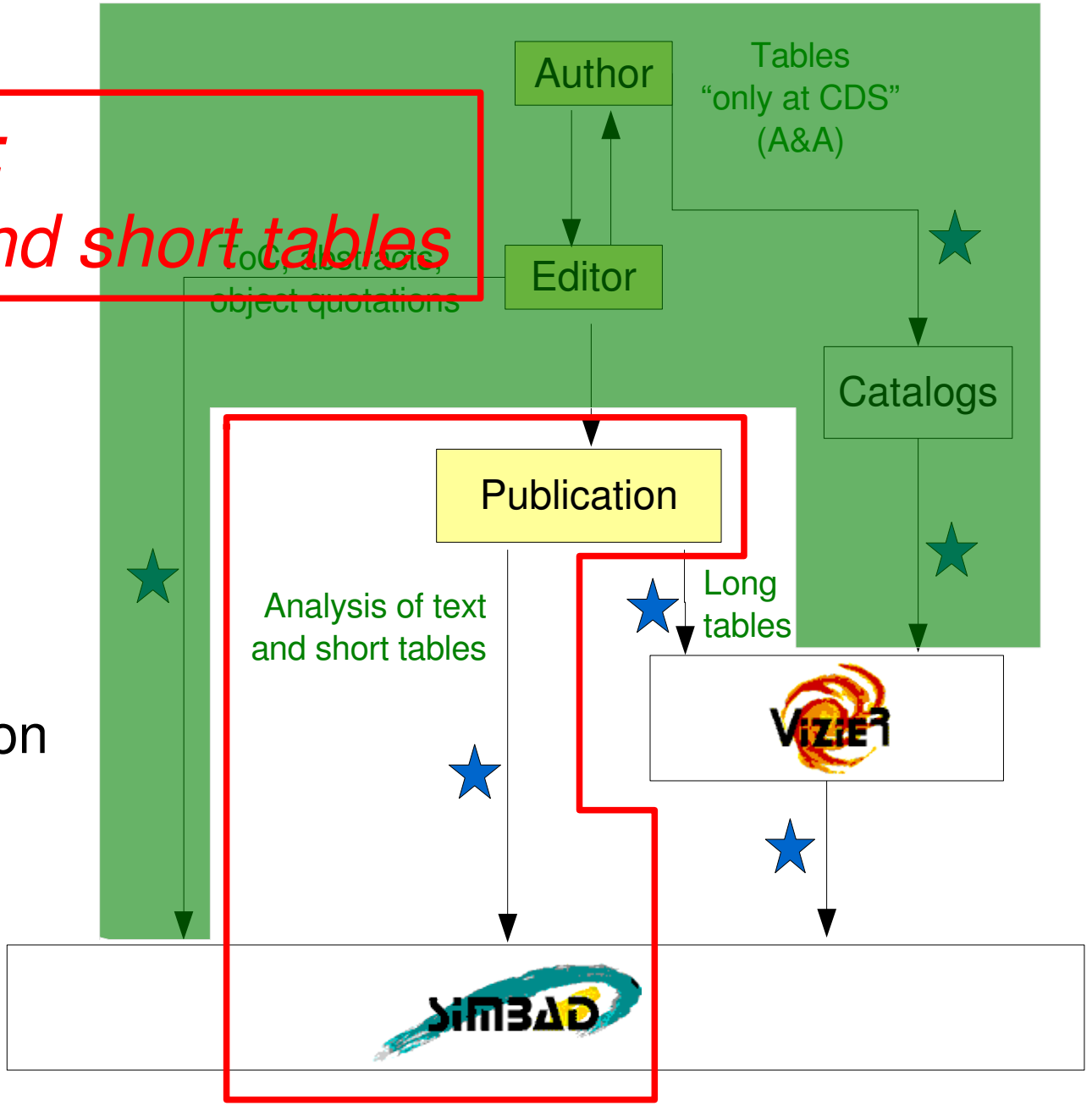


Tables of contents (2)

- Publisher
  - Tables of Contents + Abstracts + Keywords
  - Object Quotations (A&A), tagged by the authors
- CDS
  - Check/control/validation
  - SIMBAD update

How data are entered into SIMBAD (1)

Two paths:
1- Text and short tables



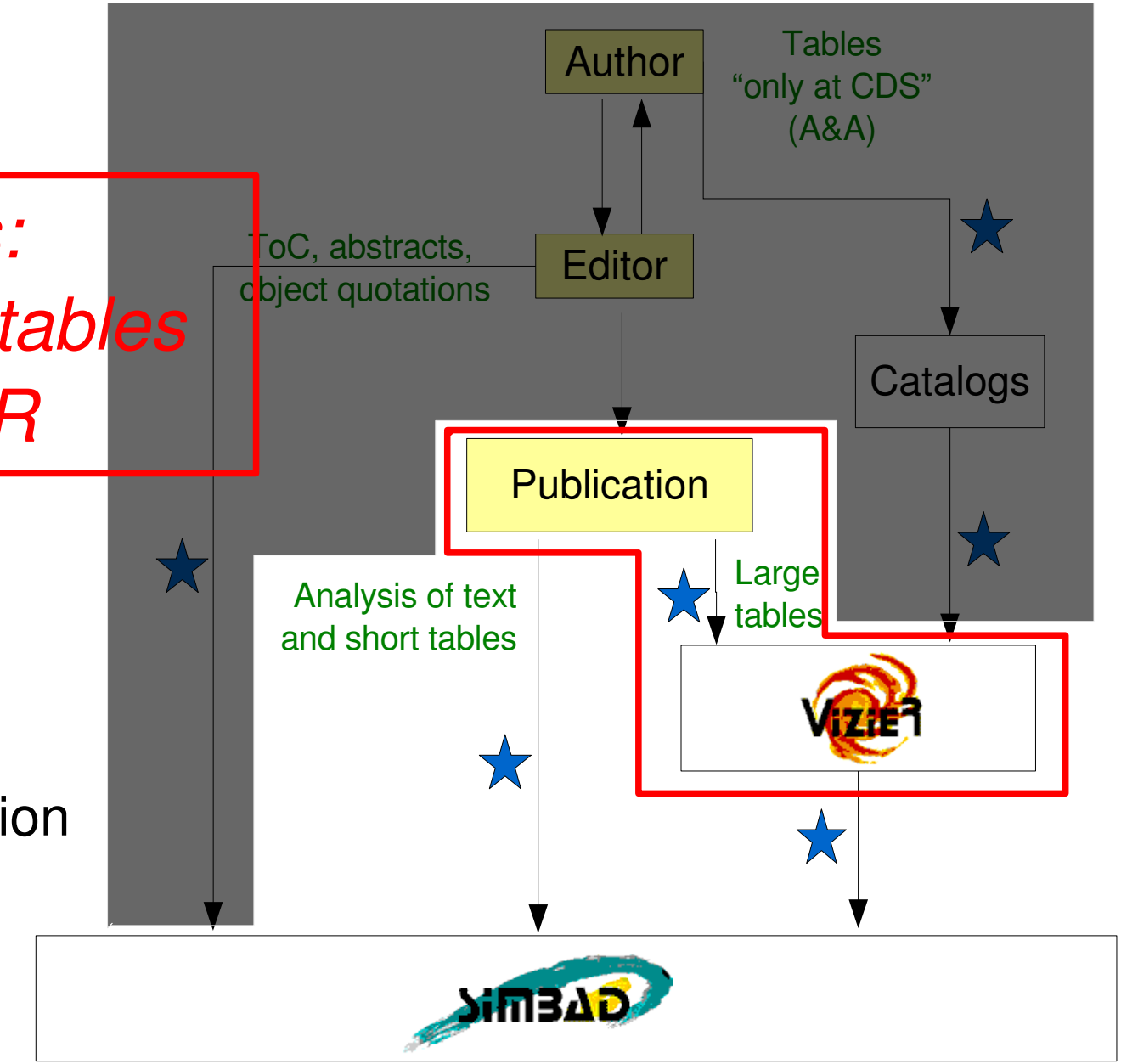
- ★ : information scientist's expertise

Detailed article analysis

- Objects in text and short tables
  - Identification of astronomical objects and related data
    - by reading the paper
      - with the help of a dedicated in-house software (DJIN, see S. Lesteven presentation) since 2008 and only for the main journals
  - SIMBAD update
    - cross identification: does the object already exist in SIMBAD (possibly under another name)?

How data are entered into SIMBAD (2)

Two paths:
2. Large tables
2.1 VizieR



- ★ : information scientist's expertise

Large tables
First into VizieR

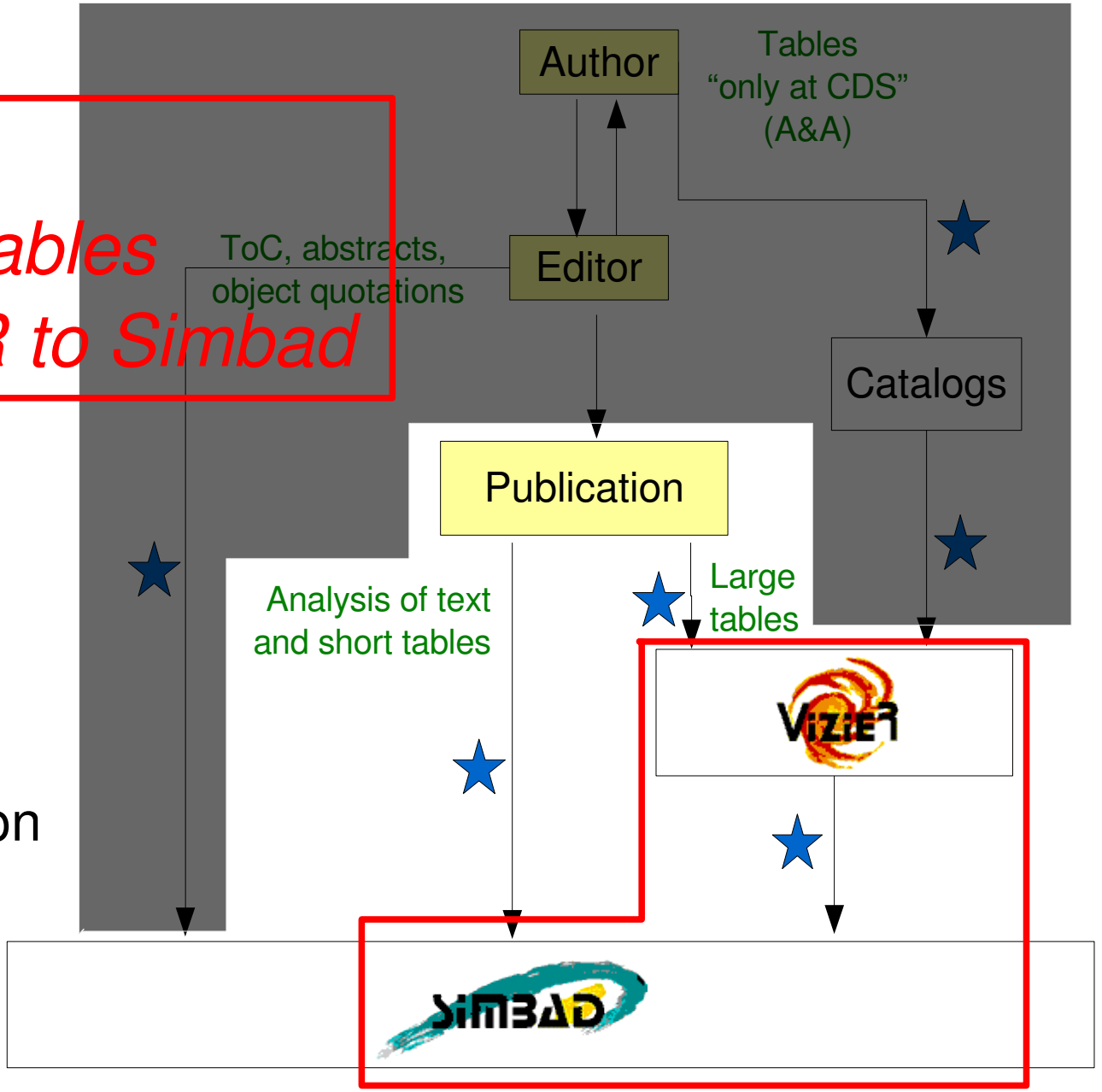
- Analysis of large tables
  - Descriptions homogenized and standardized
    - if no description file is available, one is created
  - Checks for data consistency
  - VizieR update

How data are entered into SIMBAD (3)

Two paths:

2. Large tables

2.2 VizieR to Simbad

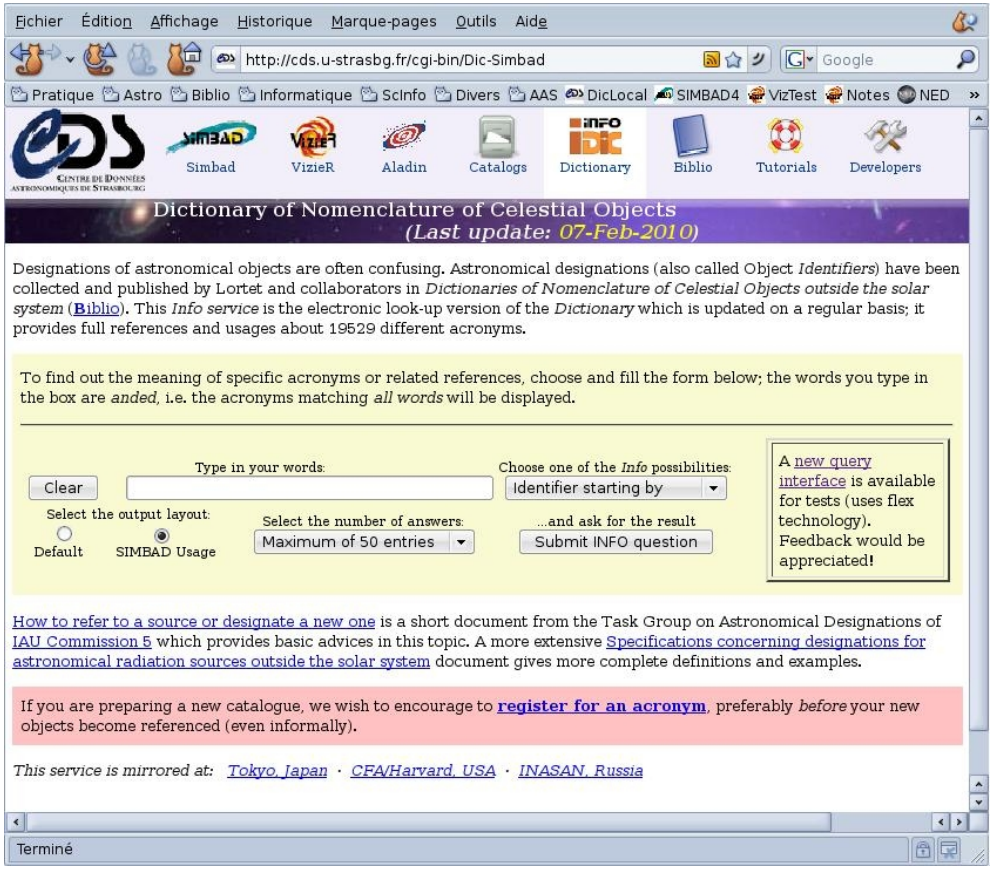


- ★ : information scientist's expertise

Large tables
From VizieR to Simbad

- Table processing
  - Identification of astronomical objects and related data
    - in-house tool “Raccord” (“*junction*”/”*join*”)
    - cross match with SIMBAD
  - SIMBAD update
  - See F. Woelfel et al. (2006, LISA V, Boston)

Our cornerstone:
The *Dictionary of Nomenclature of Celestial Objects*



- Confusing designations
- 19 000 acronyms
- IAU Commission V: register an acronym
- Essential input to DJIN
- ... It keeps me busy!

http://cds.u-strasbg.fr/cgi-bin/Dic-Simbad

What special skills do we need? (1)

- Knowledge of astronomy
  - to identify astronomical object names
  - to be familiar with the astronomical zoo (stars, binaries, cepheids, clusters of stars, planetary nebulae, interacting galaxies, voids...)
  - to recognize the parameters relevant for SIMBAD (redshifts, magnitudes, spectral types, metallicity...)
  - to follow the evolution of the discipline

What special skills do we need? (2)

- Information technology skills
  - linux environment
  - scripting and programming (in awk, perl)
- Sharing expertise
  - close interactions
  - discussion of complex cases with astronomers

Conclusion-1

- CDS: a team of complementary skills
  - information scientists
  - astronomers
  - software engineers
- Collaboration with journal editors quite effective, data and bibliography fully linked
- Widely used services with a high added-value
  - ~300,000 queries/day

Conclusion-2

- The critical elements for success
  - evolution with the scientific knowledge
  - constant interaction with astronomers
  - contents & tools: always the user's needs in mind
  - close collaboration with all others actors (journals, ADS, NED)