

Preparing for CRIS: Challenges and Opportunities for Systems Integration at Finnish Universities

Miika Nurminen (minurmin@jyu.fi), IT Services, University of Jyväskylä, Finland

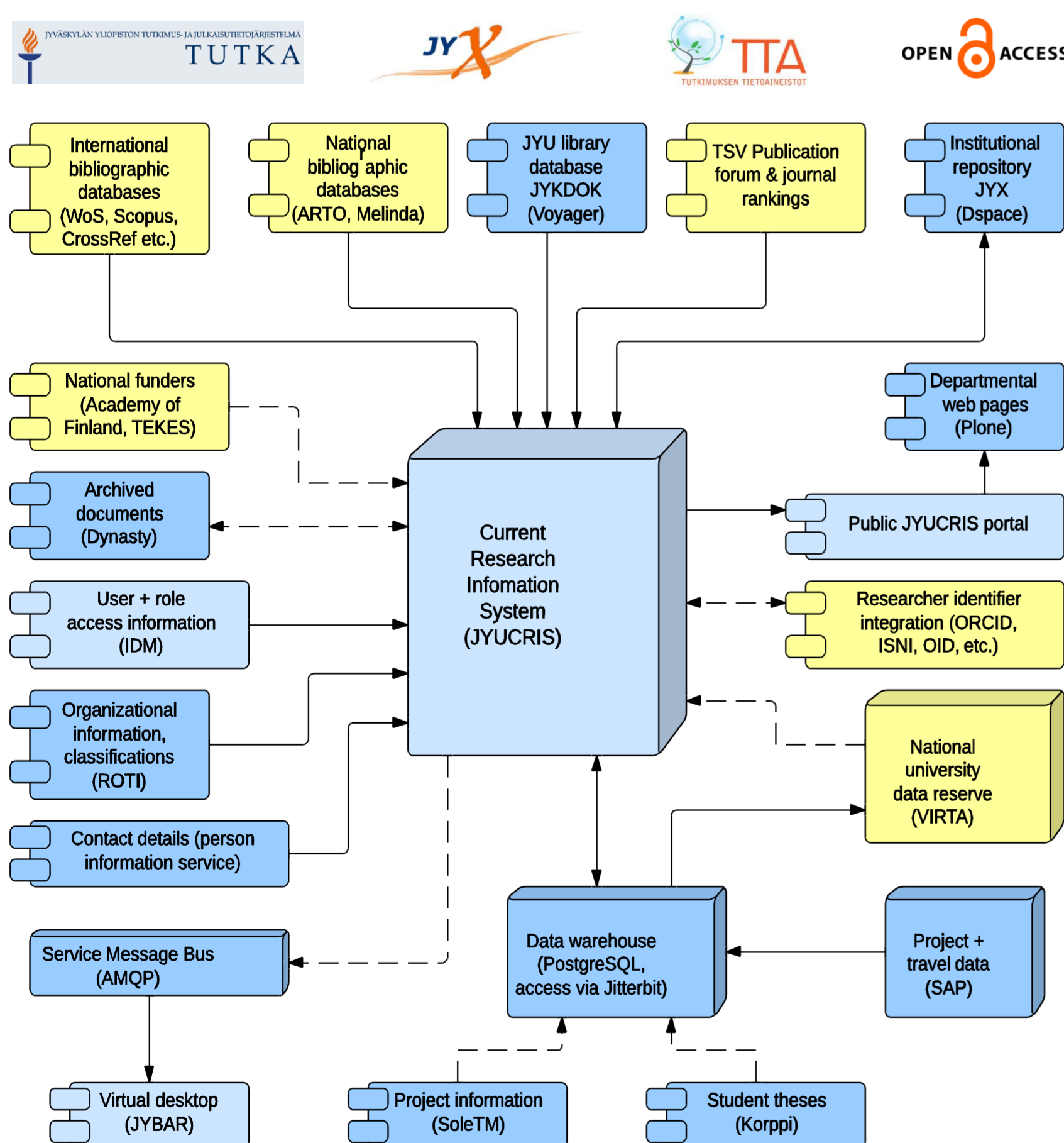
ABSTRACT

This poster presents issues identified on the attempt to integrate administrative research information with institutional repository (IR) and other systems. The observations are based on preparing for procurement of a Current Research Information System (CRIS) at the University of Jyväskylä. The CRIS will be used by various stakeholders in different organizational units, having conflicting requirements and different notions on system usage (e.g. national publication reporting, project management, researcher CVs). Determining the optimal data flow for handling publications, organizations, or financial information in different systems needs architectural consideration.

While it would be preferable to integrate CRIS and IR to a single system, there are issues with user interface, recording conventions, and selecting the data to be included that may make the task impractical. However, both systems will benefit from synchronization of selected datasets and separation of responsibilities. Even without a national publication registry, Finnish universities would benefit on services that assist on federated handling of publication data (e.g. publication forum class calculation, determining collaborative publications, transformations for data import/export). The overall publication reporting process is problematic in its current form and needs revisions at the national level.

	1. Re-engineer legacy system	2. Utilize Dspace-CRIS	3. Procure a commercial system
General customizability (data model, forms)	Yes, but needs development	Yes, partially configurable	Depends on the system (either order from the vendor, or use a configuration module)
General UI capability (e.g. dynamic search boxes)	Poor (both related to data model and UI components)	Partial (CRIS entities as authority for item metadata)	Yes
Form adaptability (fields depending on data subtypes, year-specific lists)	Yes (some "wired" to code, changes to current model need development)	Partial, needs development	Yes, mostly (adaptability to multiple classifications may depend on vendor)
Duplicate detection and unification	Partial, UI issues	Lacking (or not included in public distribution), needs development	Yes (details need usability evaluation)
Support for customized validation	Yes	Partial, needs customization	Depends on the system (either order from the vendor, or use a scripting language)
Support for specialized workflows	No ("wired" to code)	Partial (mostly related to publication submission), needs development (at least in public distribution)	Yes
Flexible affiliations and evolving organizations	Partial, needs generalization	Yes (needs customization, possible UI issues)	Yes (details need usability evaluation)
Publication metadata import	MARCXML import available, WoS/Scopus import not implemented	Needs development, API available	WoS/Scopus available, MARCXML import needs development
Support for managing project information (funders, documents)	Partial (existing version needs development)	Lacking (esp. on complex workflows and project-specific document management)	Yes
Reporting functionality	Yes	Partial, needs development or external applications	Yes
Fine-grained user privileges	Yes (existing application server privileges + customization)	Partial (mostly entity-level, may need customization)	Yes
Long-term support	Considerable risk (limited number of developers, legacy code)	Yes (active open source project), however project goals may differ from organizational goals	Yes, mostly (both major vendors have multiple installations - however, both have also been acquired by a larger company)
Data migration cost	None	Combining research information with existing IR data requires extensive matching	Data from legacy system needs cleaning and transformation
Procurement cost	Minor (internal development)	Medium (some consulting probably needed)	Major
General maintenance cost	Major (app server and multiple components should be upgraded)	Major (could be combined with IR development), needs to be "synced" with DSpace-CRIS project	Medium (local data integration + yearly maintenance + additional development)

Highlights from the preliminary evaluation of different implementation options for the new CRIS (including potential re-engineering of the legacy system TUTKA). The evaluation was carried out in fall 2013 and included DSpace-CRIS 3.2.0 beta, Pure 4.16, and Converis 5.1.



Planned integrations (=subject to change) for the future CRIS. Yellow boxes denote components or systems external to JYU, blue ones are developed or otherwise managed internally (SAP is used collectively by multiple Finnish universities). Solid lines denote automated data transfers, dotted lines depict optional or manual integrations.

National context

Since 2011, the focus in national publication reporting has shifted from summary statistics to full publication metadata and ranked publication channels (=Publication forum). This has resulted to complex classification rules (e.g. determining whether an edited book is a "report" or "scientific book"; ambiguous restrictions with conference articles; omission of research contributions in non-traditional publication channels).

The publication reporting process as a whole is problematic because of its distributed (the original idea to adopt a national publication registry was abandoned), one-way nature: for collaborative publications, multiple universities end up reporting the same data that is checked, combined, and possibly sent back to original universities in case of conflicts. After the data is reviewed and assigned a rank, erroneous data can no longer be corrected. Finnish universities would benefit on services that assist on federated handling of publication data, utilizing national service bus for basic infrastructure.

- One step towards the right direction has been extending the national article registry ARTO, provided that the data is usable in time and easily harvestable to existing systems, with appropriately marked affiliations.
- The publication forum has issues, since the exact way to calculate the ranking for varying publication types and spellings can be ambiguous. There should be a web service API to calculate the ranking and submit publication channels for evaluation with respect to exact publication metadata - used by both the ministry and the universities.
- In the future, researcher identifiers may be used with publication reporting. This would necessitate a service used to automatically check and generate identifiers (e.g. ORCID), perhaps based on national student identifiers (OID) which in principle already provide a mechanism to preserve identity even if a researcher changes institution.
- The ability to automatically transfer full (=conforming to national requirements, e.g. with a CERIF-like model) publication metadata between university-level systems - and a service to automate these transfers - is desperately needed. This would enable entering publication data for national collaborative publications incrementally, utilizing the data from different organizations. A centralized "publication messaging" service would also streamline the overall reporting process.

CRIS development efforts at the JYU

University of Jyväskylä has been using an in-house developed research information system TUTKA since 2003. The system has been under continuous small-scale development. Due to increased internal (esp. research project management) and external (=national publication reporting) requirements, usability problems, obsoleted basic components, and the need for new integrations (e.g. WoS/Scopus import), the university has decided about CRIS renewal project with commercial system as the primary option.

- Spring 2013: TUTKA group¹ outlined the scope (publications, projects with additional funding, international mobility, essential scientific prizes) and primary requirements and for the new system.
- Fall 2013: Evaluation of different implementation options, sketching the architecture.
- Spring 2014: JYU Library started centralized recording of publications. The goal is to ease researchers' activities and to promote parallel publishing to JYX. The same principle (=centrally recorded, validated data) will be applied with the new CRIS and extended to other types of entities as well.
- Now: preparation for tendering, detailed requirements, processes, and data model.
- Fall 2014 (expected): tendering process, cleaning up data from the legacy system.
- 2015 (expected): Legacy data transfer, incremental adoption of the new CRIS.

¹ <https://www.jyu.fi/hallinto/tyoryhmat/tutkaryhmat/en/>

