

Family History Technology: A Survey of 10 Hard Problems

Dr. Doran Wilde

Brigham Young University

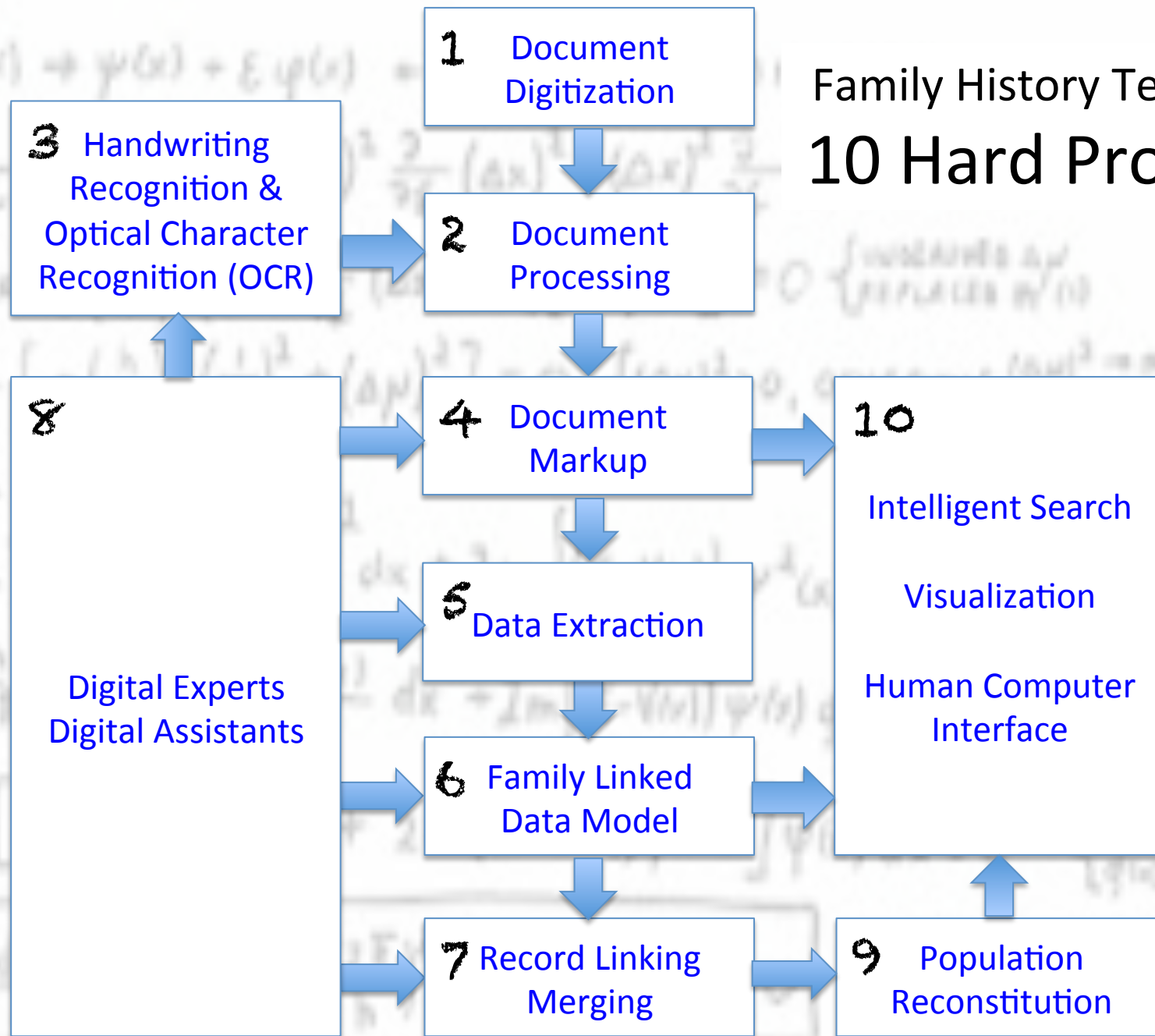
Dept. Electrical and Computer Engineering

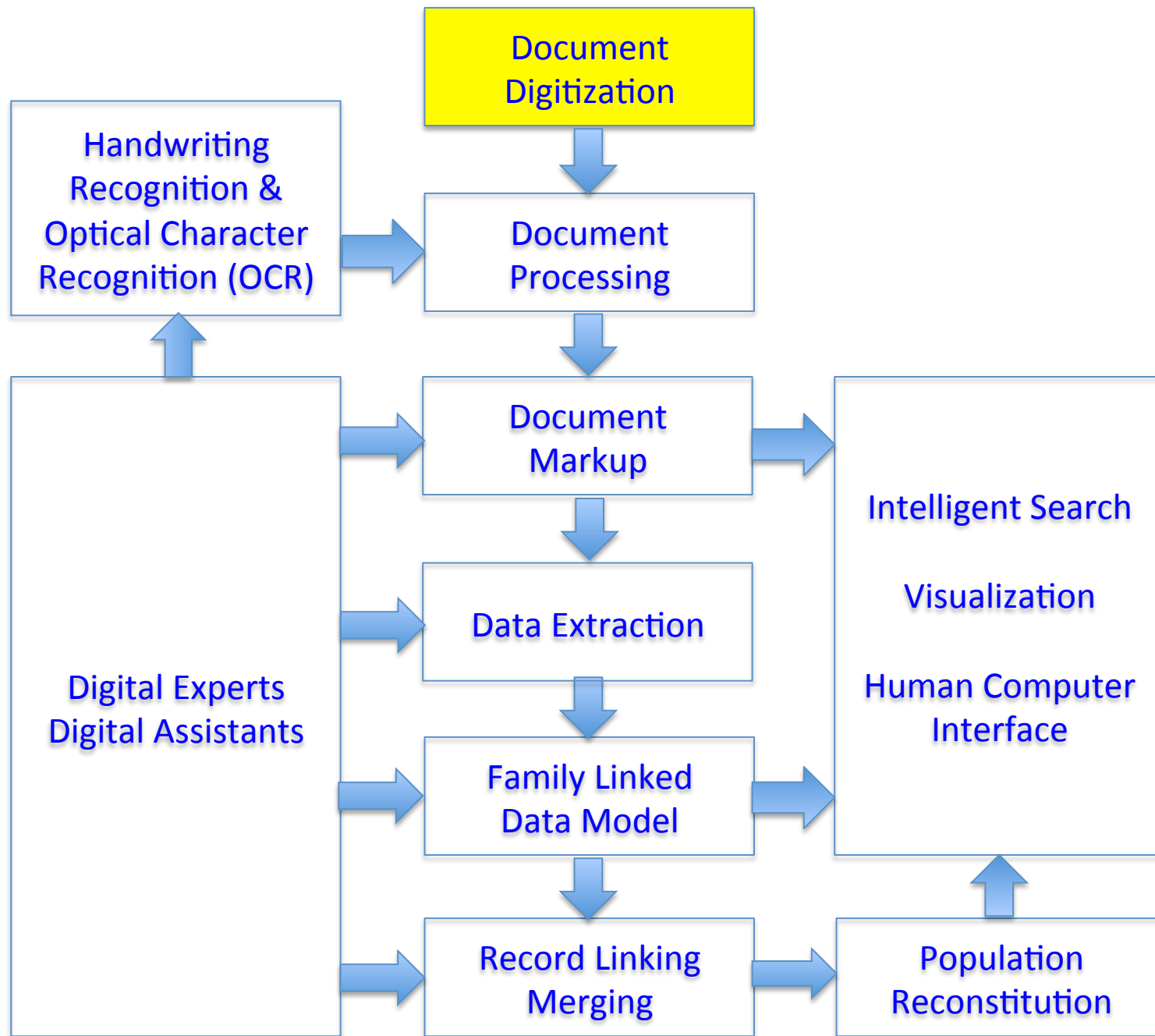
Family History Technology Workshop 2014

Decomposing a Really Really Hard Problem

- The goal is to build a common lineage-linked family tree for the whole human race based upon available records.
- Decompose this seemingly impossible challenge into 10 hard problems.
- Allows researchers to focus on a part of the overall problem.
- Illustrated by a running example.

Family History Technology 10 Hard Problems

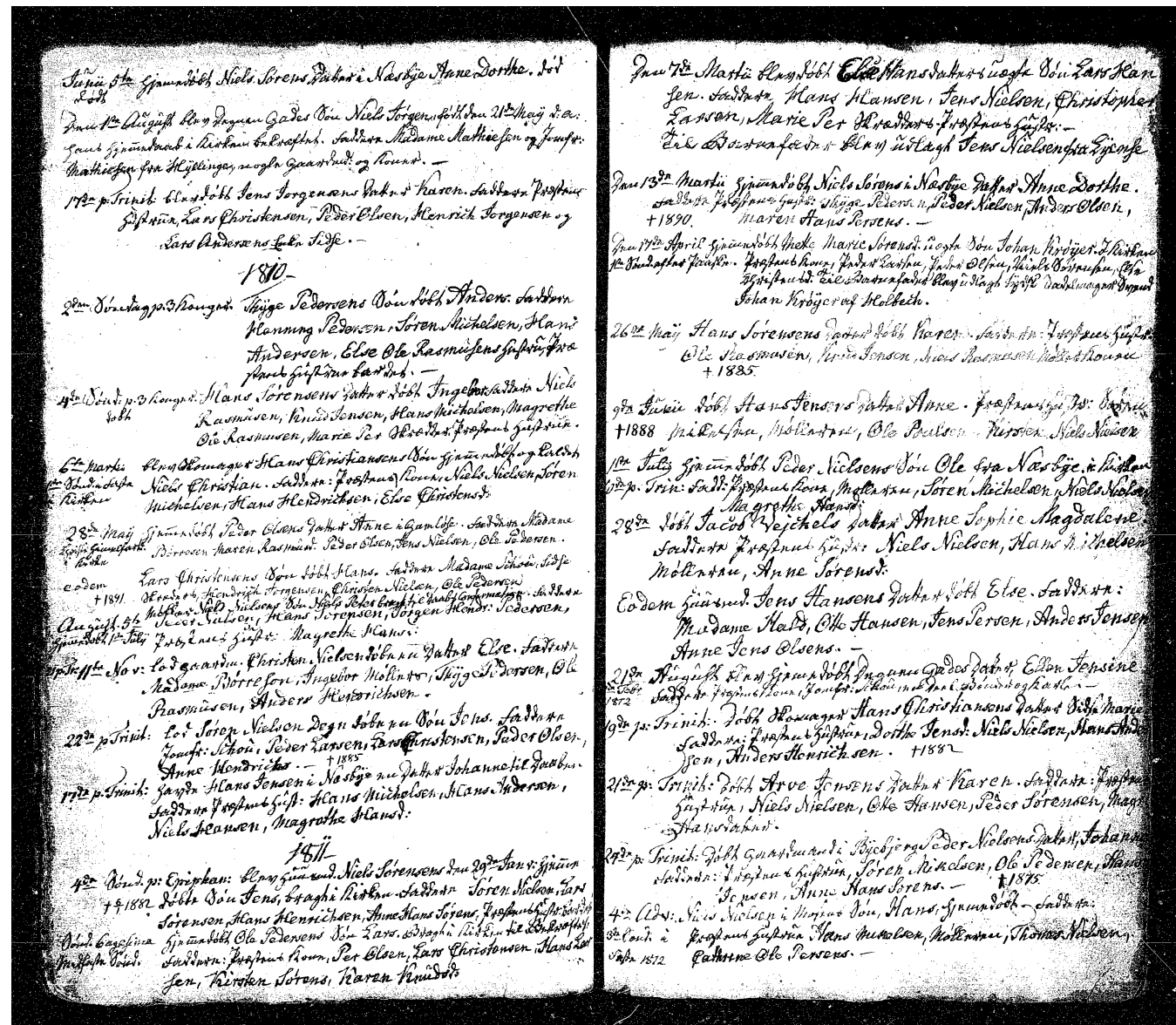


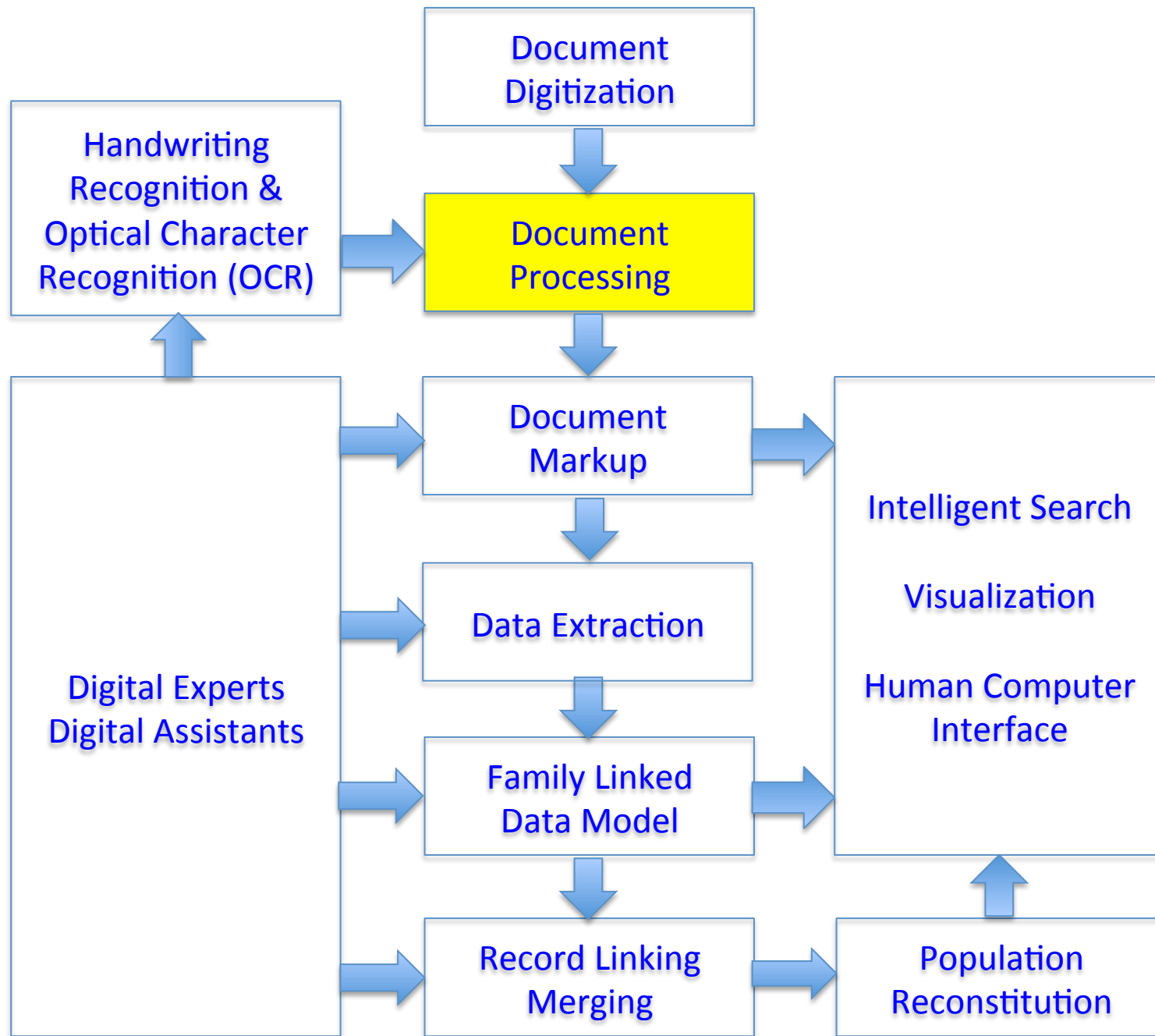


Document Digitization and Preservation



Preservation,
Accessibility, and
Enhancement of
Records

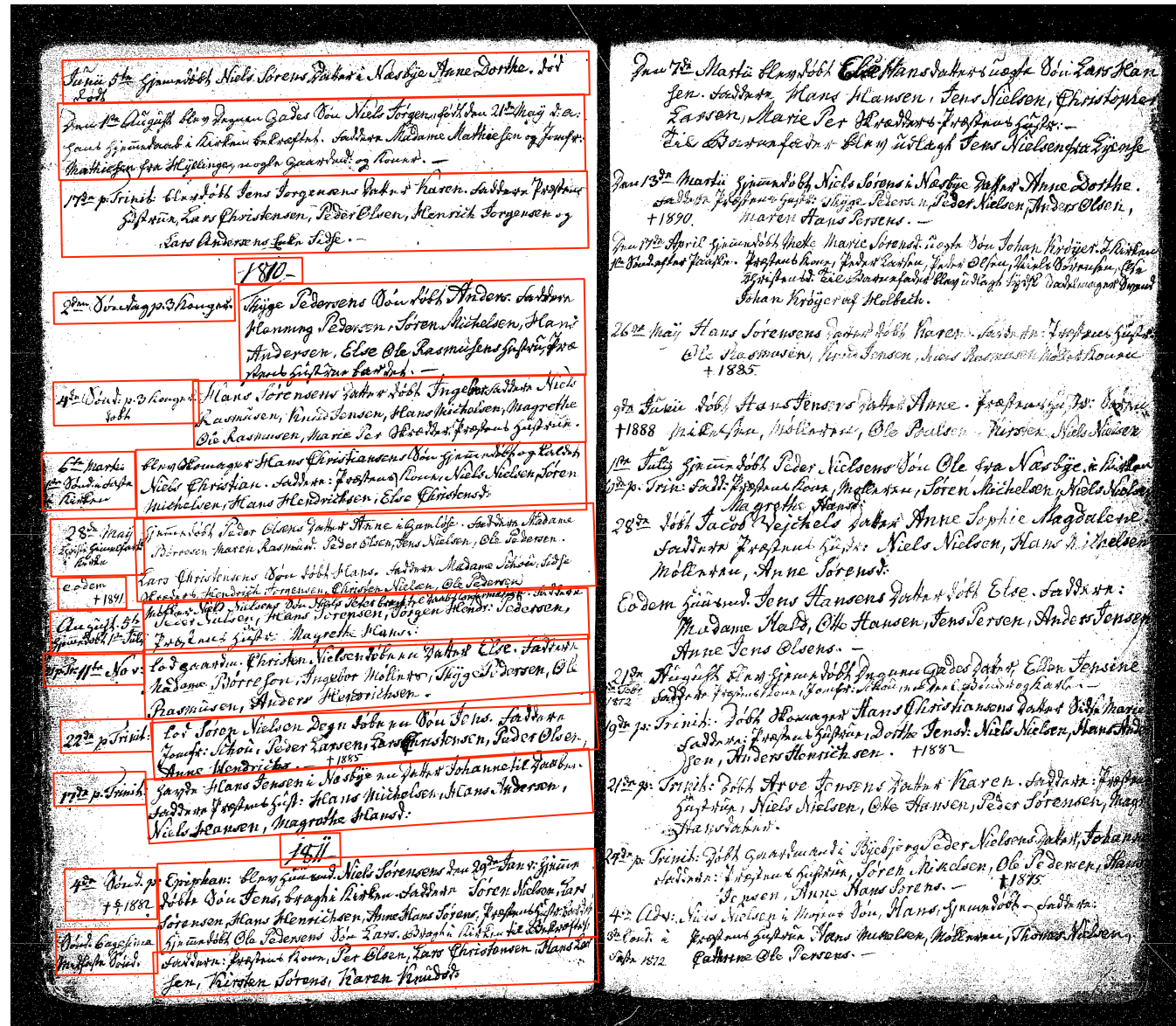


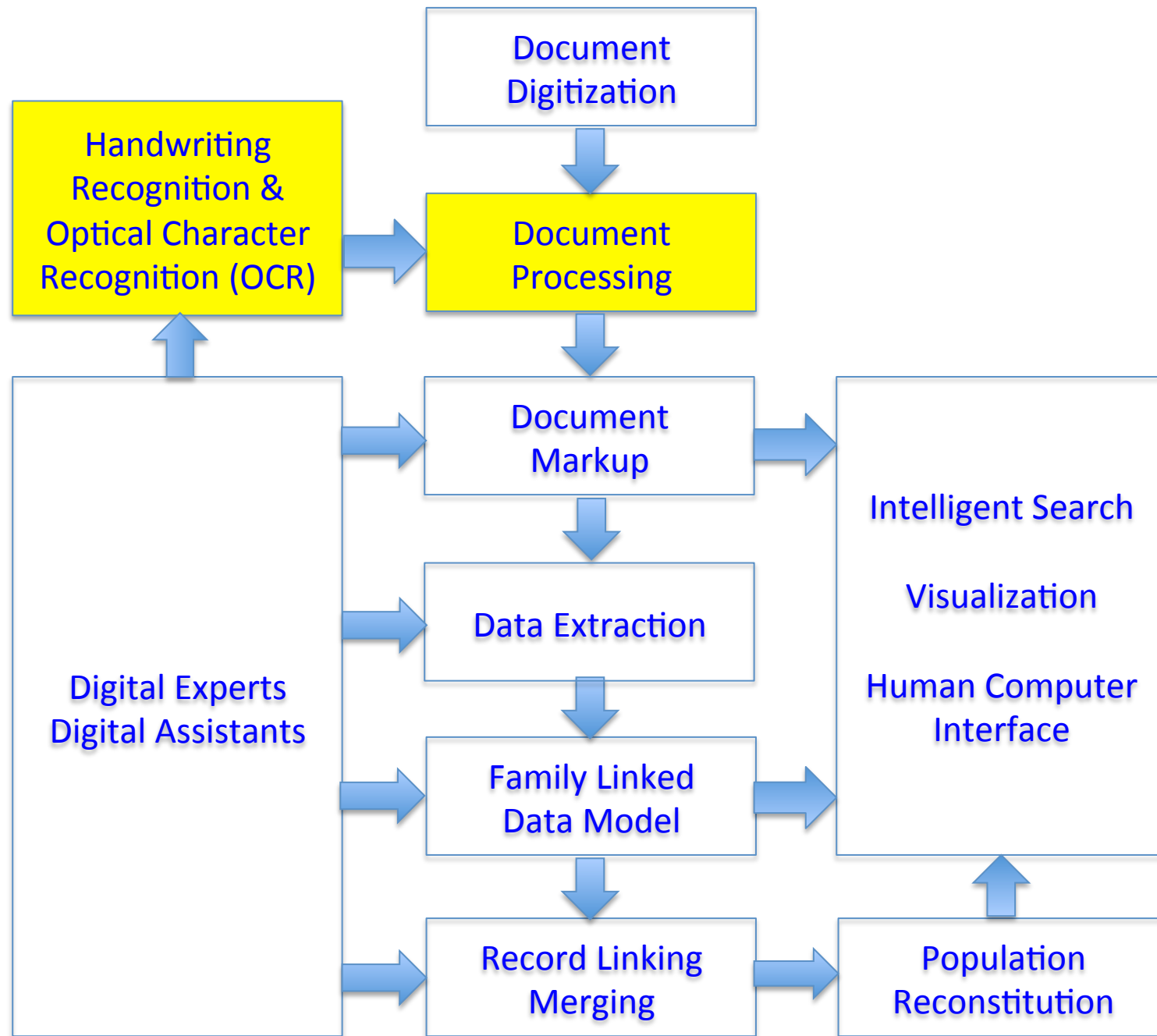


Document Processing

Spatially parse
the page into
blocks to give
context and
meaning to
what is
written.

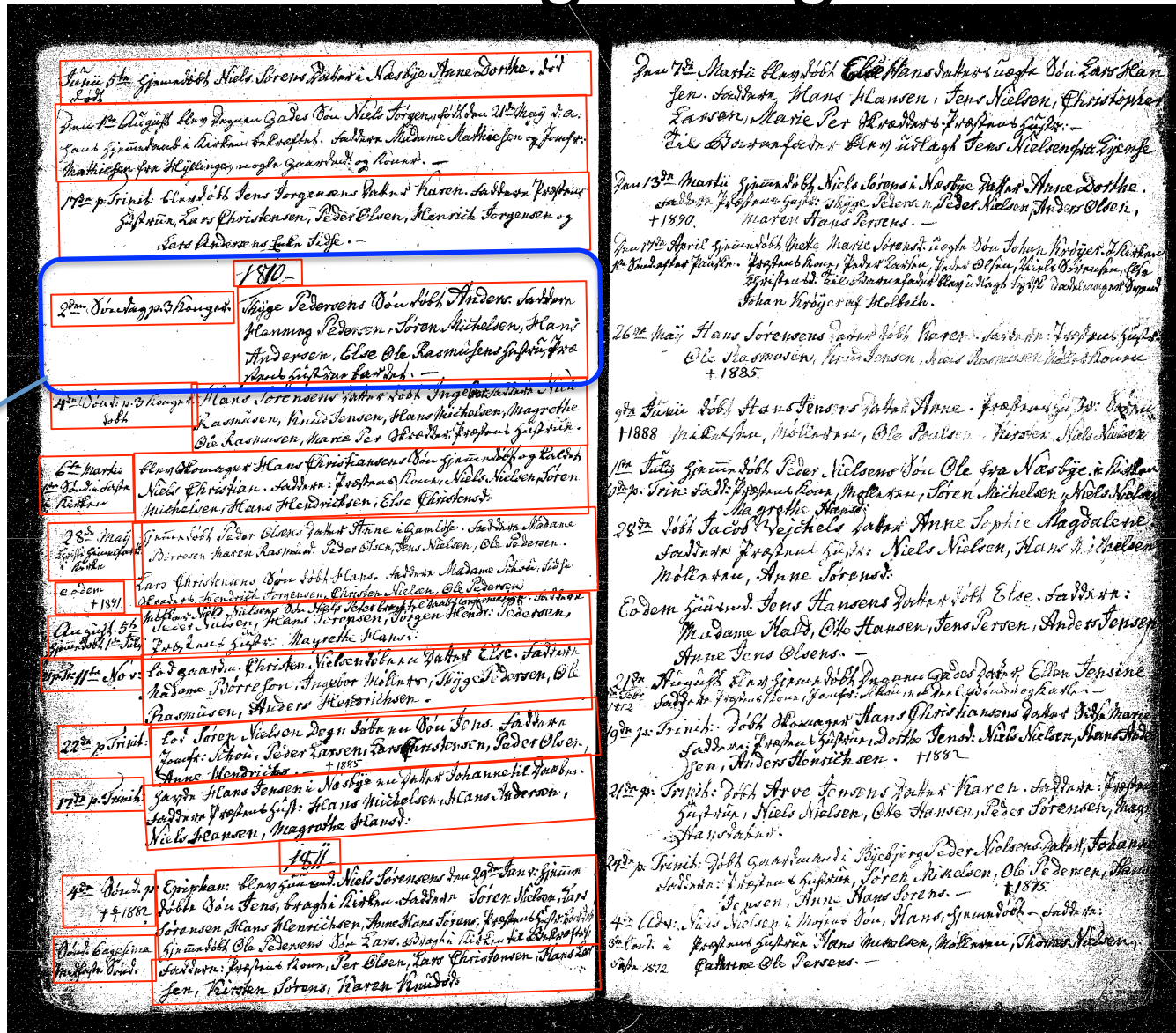
Zoning.





Document Processing and Handwriting Recognition

Lines passed from image to handwriting recognizer, and collates the results back into an annotated transcript of the record.



Document Processing and Handwriting Recognition

1810.
2^{den} Søndag p. 3 Konger. Thyge Pedersens Son dobt Anders. Faddern
Henning Pedersen, Søren Michelsen, Hans
Andersen, Else Ole Rasmusens hustru, Præ
stens hustrue bar det. —

[Annotated transcript]

[Christening Record]

[Source: Church records, Oro, Frederiksborg, Denmark]

[Image 112]

[Heading] 1810

[Margin Note] 2^{den} Søndag p. 3 Konger.

[Body]

[Line 1] Thyge Pedersens Son dobt Anders. Faddern

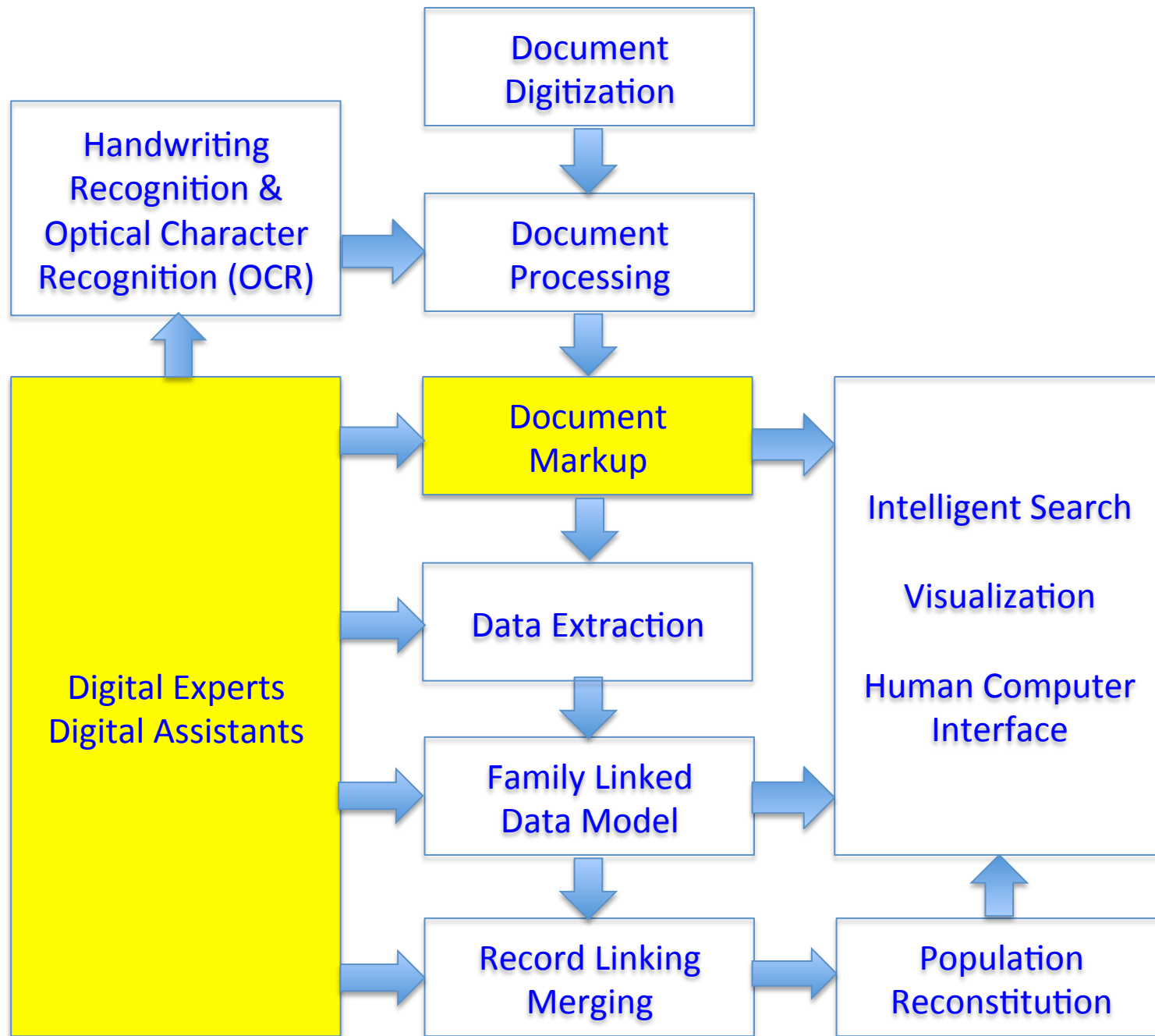
[Line 2] Henning Pedersen, Søren Michelsen, Hans

[Line 3] Andersen, Else Ole Rasmusens hustru, Præ

[Line 4] stens hustrue bar det. —

[Annotations are added or inferred by the document processor with human input.]

A **transcript** is a **faithful** copy that preserves the characteristics of the original as much as possible, making a useful copy for further analysis and research.



Document Markup

[Source: Church records, Oro, Frederiksborg, Denmark]

(Danish)

[Image 112]

[Heading] 1810

[Margin Note] 2^{den} Sondag p. 3 Konger.

[Body]

[Line 1] Thyge Pedersens Son dobt Anders. Faddern

[Line 2] Henning Pedersen, Soren Michelsen, Hans

[Line 3] Andersen, Else Ole Rasmusens hustru, Præ

[Line 4] stens hustrue bar det. —

Categorize text into semantic elements: names, dates, places, etc. Adds markup to the document to convey this information.

Uses Date Authority
Uses Name Authority
Uses Place Authority

[Source: Church records, Oro, Frederiksborg, Denmark]

(English Translation)

[Image 112]

[Heading] 1810

[Margin Note] 2nd Sunday past 3 Kings.

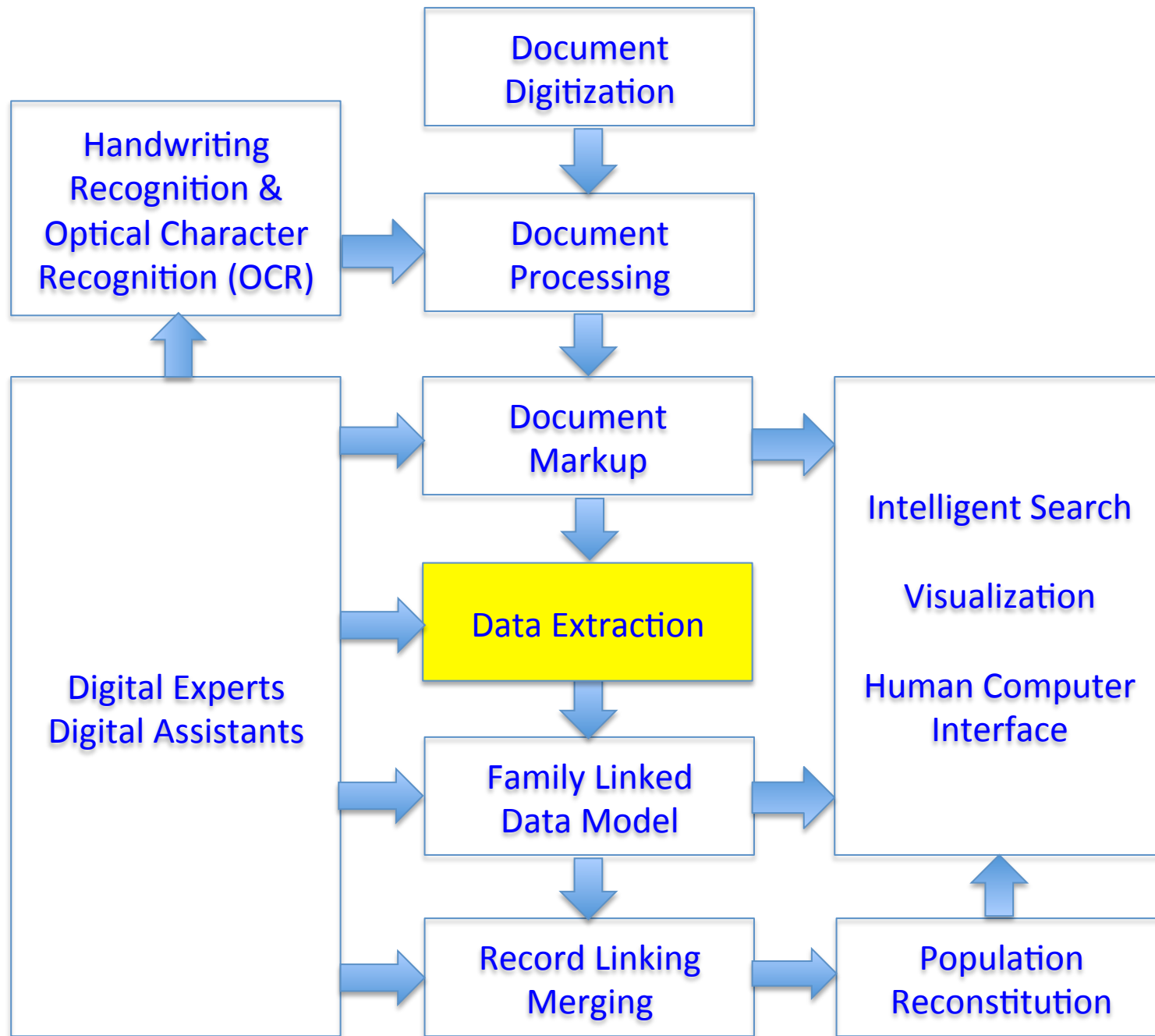
[Body]

[Line 1] Thyges Pedersen's Son christened Anders. Witnesses

[Line 2] Henning Pedersen, Soren Michelsen, Hans

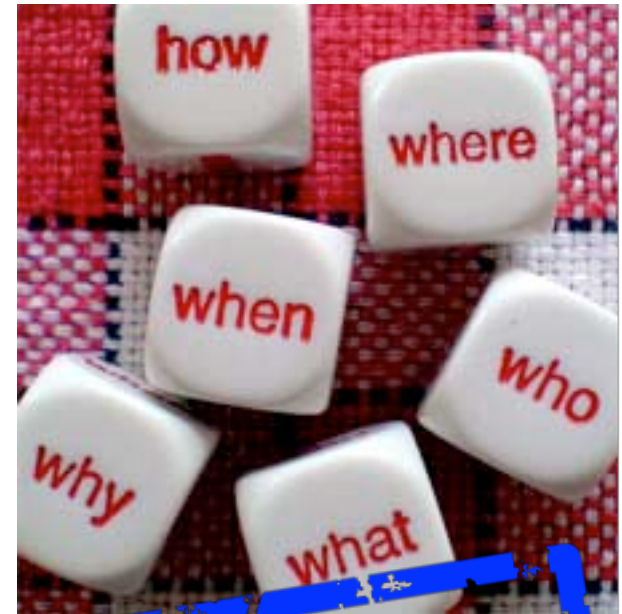
[Line 3] Andersen, Else Ole Rasmusen's wife, the Priest's

[Line 4] wife carried the child. —



Facts

- Also called assertions (facts with associated confidence ratings).
- Answer the questions:
 - what (fact type)
 - who/role (might be multiple persons)
 - when (date)
 - where (place)

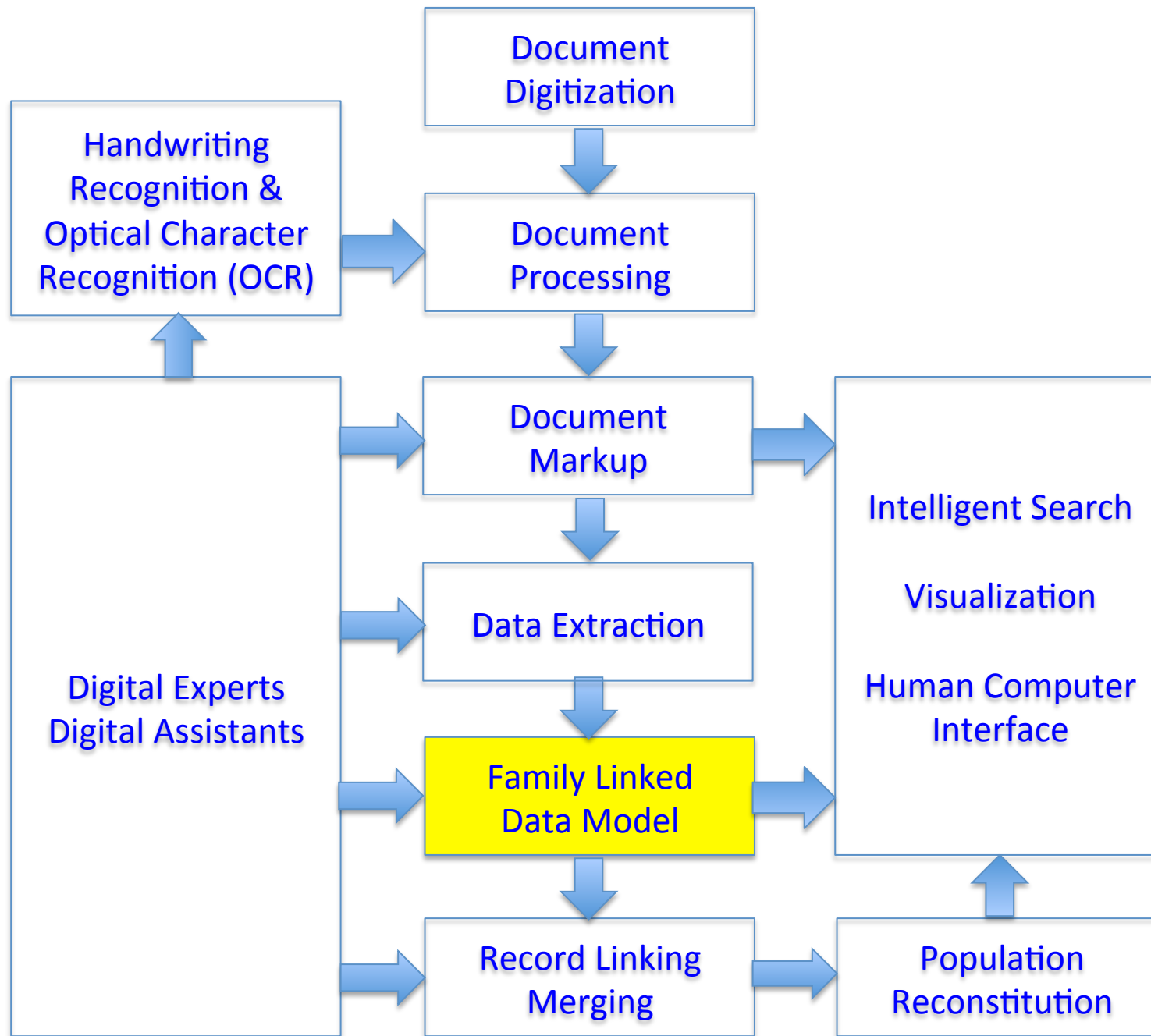


Data Extraction

Inferred Facts:

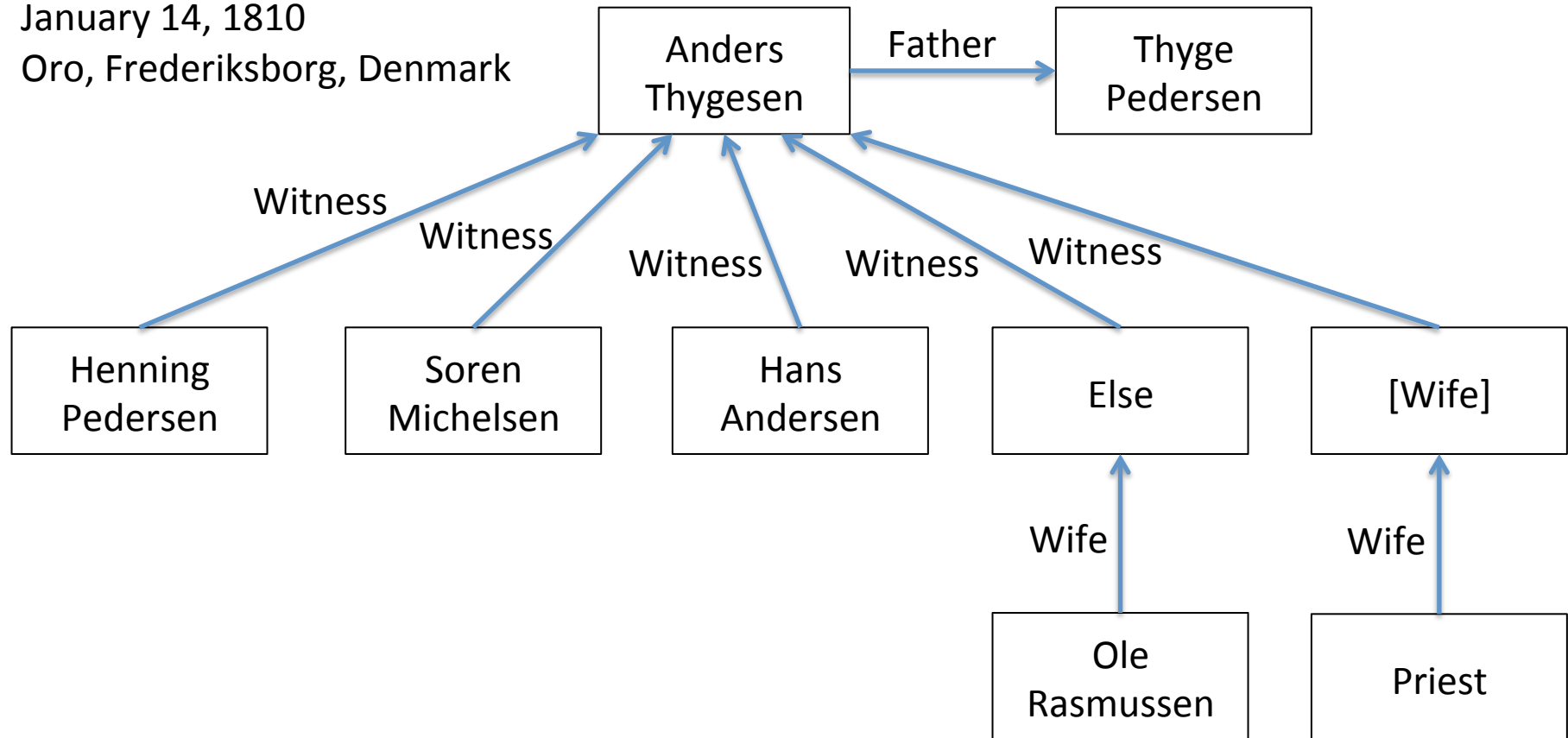
- Source Citation: Oro, Fredericksborg, Denmark Church Records, Christenings, Image 112 [Document Processor]
- Location is Oro, Fredericksborg, Denmark [Place Authority]
- 3 Kings Day is January 6. [Date Authority]
- 2nd Sunday past 3 Kings Day in 1810 is January 14, 1810. [Date Authority]
- Anders Thygesen was christened
- Thyges Pedersen was the father of Anders Thygesen
- Henning Pedersen was a witness to the christening of Anders Thygesen
- Soren Michelsen was a witness to the christening of Anders Thygesen
- Hans Andersen was a witness to the christening of Anders Thygesen
- Else was a witness to the christening of Anders Thygesen
- Else was the wife of Ole Rasmusen
- The Priest's wife was a witness to the christening of Anders Thygesen

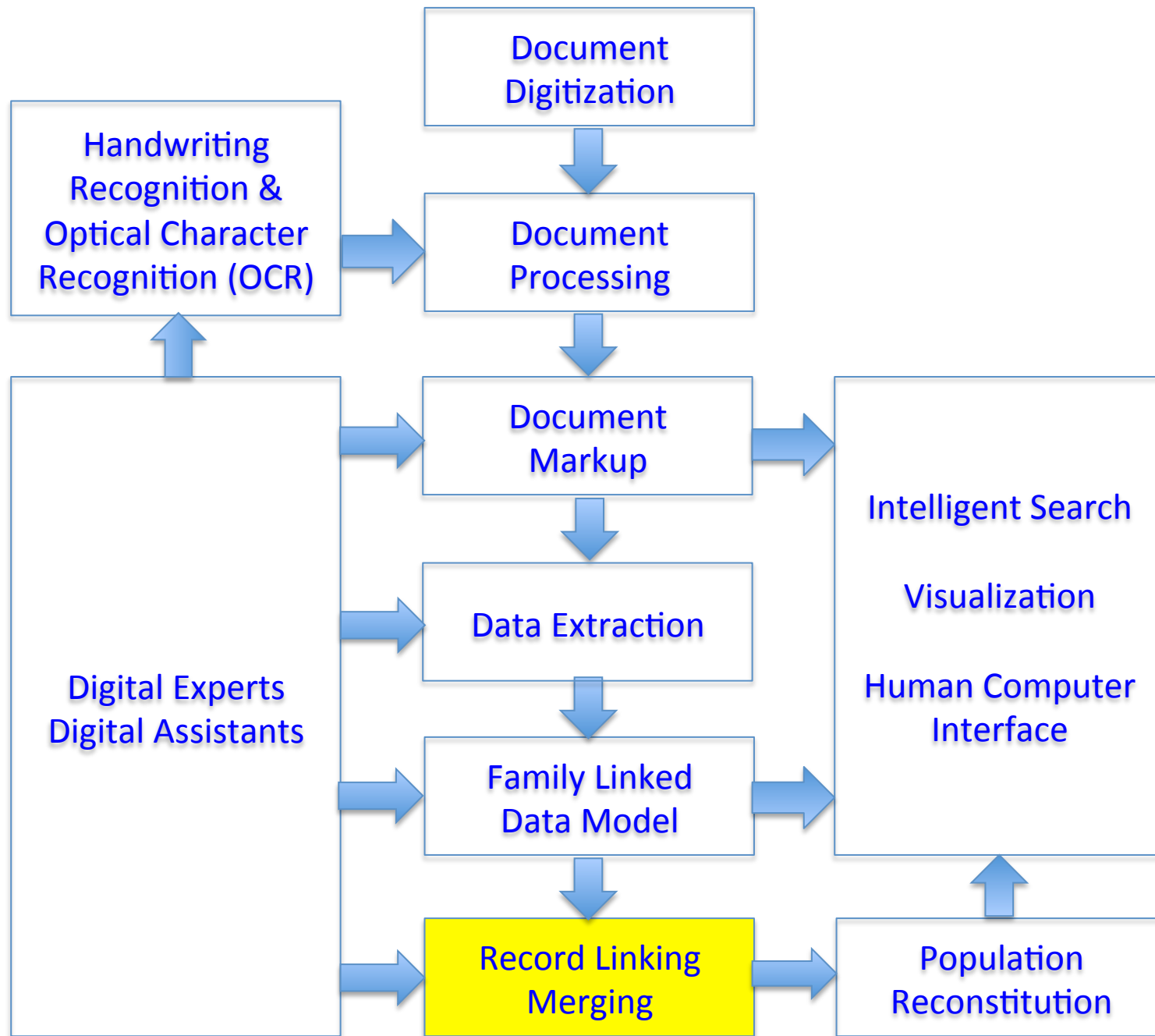
(12 Facts from this simple Christening record.)



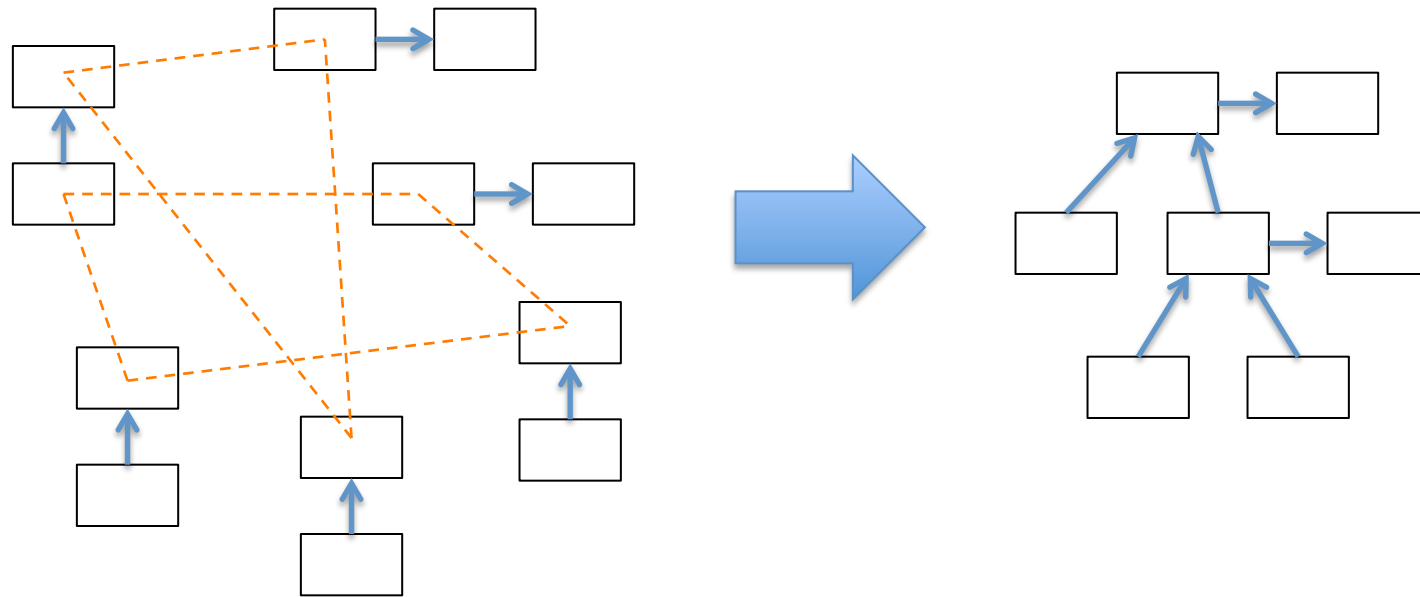
Family Linked Data Model

Christening Event
January 14, 1810
Oro, Frederiksborg, Denmark



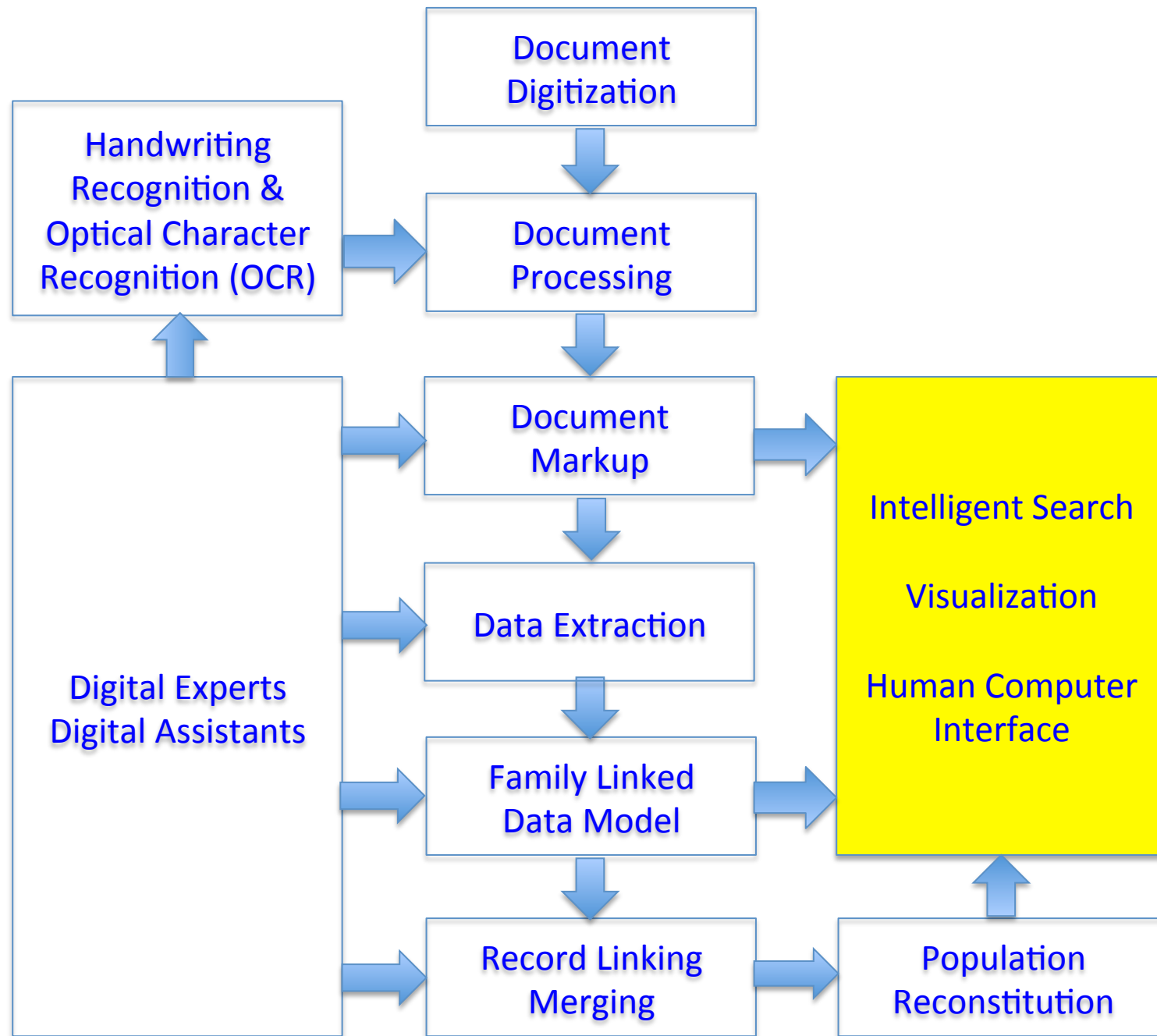


Record Linkage/Merging

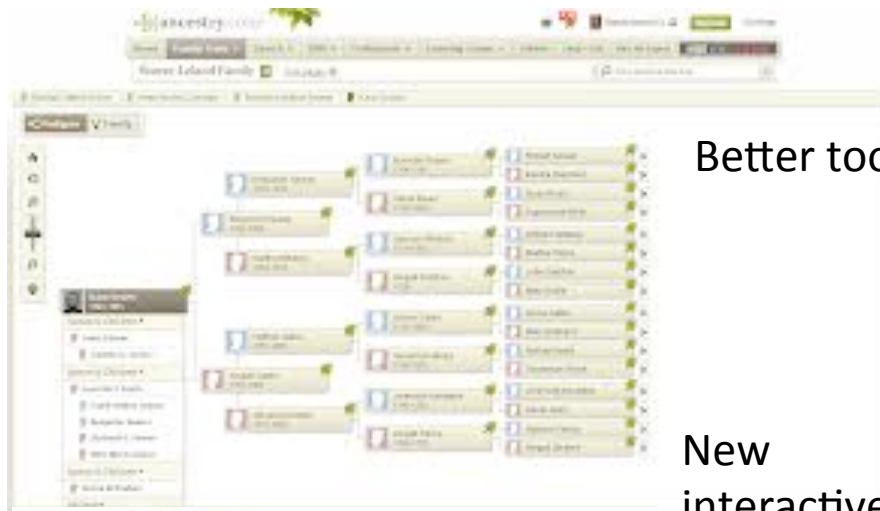


Record linking determines when two or more different records refer to the same individual.

Duplicates are merged together to build lineage-linked families.

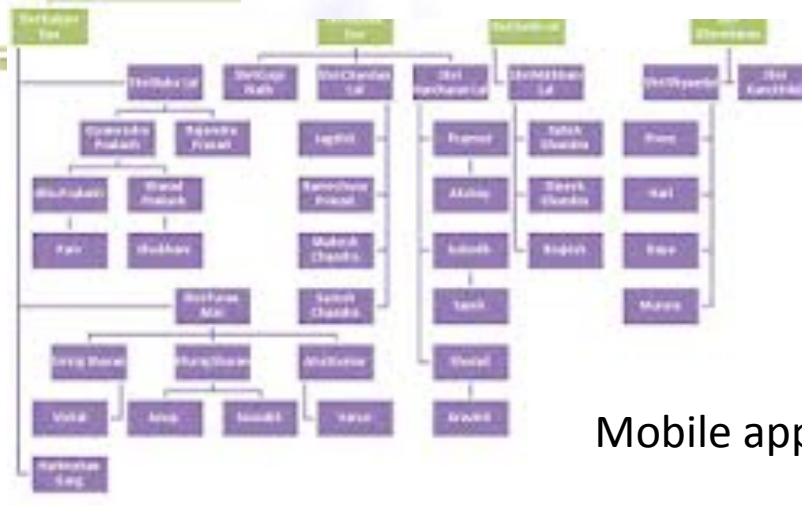
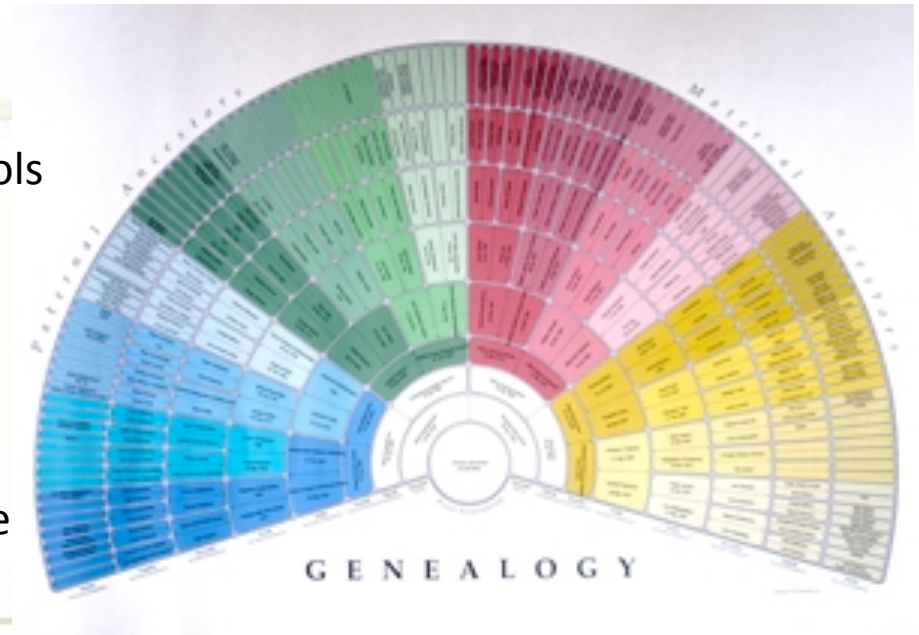


Search, Visualization & Computer-Human Interface

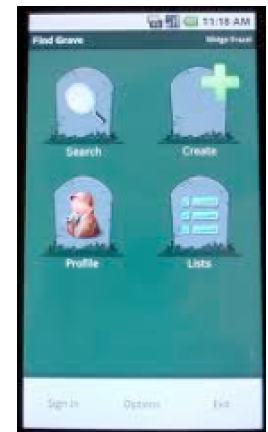


Better tools

New interactive charts.



Mobile apps



Summary

- I have surveyed 10 problems that are on the critical path to a revolutionary break-through in the way we do genealogical research.
- There are undoubtedly other important problems that impede family history that haven't been mentioned or haven't yet been identified.
- The Family History Technology Workshop is held to discuss these kinds of problems.
- The hope is that technology can put the ultimate goal of assembling a family tree for all people within reach.